DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

PROBABILITY FEEDBACK IN A RECURSIVE SYSTEM OF LOGIT MODELS: ESTIMATION

Quang H. Vuong

SOCIAL SCIENCE WORKING PAPER 444

July 1982

ABSTRACT

Some estimation procedures are considered for the recursive system with probability feedbacks introduced in Q. H. Vuong (1982b) for the case in which the probability models are logit models. The asymptotic distributions of the two-step estimator, of the estimator obtained at each iteration of a natural iterative sequential procedure, and of the estimator obtained at each iteration of an efficient iterative sequential procedure are derived. It is shown that, upon convergence, this latter procedure produces an asymptotically efficient estimator.

PROBABILITY FEEDBACK IN A RECURSIVE SYSTEM OF LOGIT MODELS: ESTIMATION

By Quang H. Vuong [1]
California Institute of Technology

In a previous paper (Q. H. Vuong (1982b), we introduced a general model defined by a recursive system of probability models in which the conditional probabilities of posterior equations are fed back into anterior equations. Specifically, if A and B are two endogenous qualitative variables, then a recursive system with probability feedback for these two qualitative variables is defined by a pair of probability models (one for the conditional distribution of, say, A given B, and one for the distribution of B) in which the probability that B takes on any particular value depends among other things on some of (or possibly all) the conditional probabilities for A given B. Such a recursive model was motivated by various examples. In particular, it was shown that the multivariate logit model (see e.g., M. Nerlove and S. J. Press (1973, 1976)) and the constrained nested logit model (see, e.g., D. McFadden (1981)) are special cases of the recursive model with probability feedback. Two examples in a game setting (a game against Nature, and a Stackelberg game under uncertainty) illustrated, however, the more general formulation.

In the present paper, we are interested in the estimation of such a model for the case in which the probability models that constitute the recursive system are logit models. To simplify the discussion, the next results are derived within a two-equation framework. The results can, however, be extended to the case of more than two endogenous qualitative variables.

Let I and J be the number of categories of the two qualitative variables A and B.[2] Let t denote the t-th individual of a sample of size T. The probability that the t-th individual "chooses" the alternative j of B is $\Pr(B = j|t)$, while the probability that that individual "chooses" the alternative i of A given that he has chosen j for B is $\Pr(A|B = j,t)$. For any t, let $\Pr(A|B,t)$ and $\Pr(B|t)$ be the vector of conditional probabilities $\Pr(A = i|B = j,t)$ and the vector of probabilities $\Pr(B = j|t)$. Then the recursive system with probability feedback is defined by:

for any $t = 1,\ldots,T$, $i = 1,\ldots,I$, $j = 1,\ldots,J$:

$$\log \Pr(A = i|B = j,t) = \mu_{jt} + z'_{ijt}\alpha \tag{1}$$

$$\log \Pr(B = j|t) = \mu_t + z'_{jt}\beta \tag{2}$$

where

$$z_{jt} = z_{jt}(\Pr(A|B,1),\ldots,\Pr(A|B,T)) = z_{jt}(\alpha) \tag{3}$$

$$\mu_{jt} = -\log\left[\sum_{i=1}^{I} \exp(z'_{ijt}\alpha)\right] \tag{4}$$

$$\mu_t = -\log\left[\sum_{j=1}^{J} \exp(z'_{jt}\beta)\right] \tag{5}$$

The functions $z_{jt}(.)$'s are assumed to be known and twice continuously differentiable.[3] The parameter vectors α and β are unknown and their sizes are respectively a and b. The vectors z_{ijt} and z_{jt} are commensurate with α and β, and are interpreted as in standard logit analysis (see D. McFadden (1974,1981)). The vector z_{ijt} embodies observed characteristics of the i-th alternative of A and of the j-th alternative of B, as well as observed characteristics of the t-th individual. Hence the vector z_{ijt} is observed. On the other hand, the vector z_{jt}, which depends on observed characteristics of the alternative j and on observed characteristics of the t-th individual is not observed since that vector depends on the unobserved conditional distributions $Pr(A|B,t)$, $t = 1,\dots,T$, and hence on the unknown parameter α.

Various methods for estimating the model (1) - (5) are studied. First, we consider the two-step (sequential) estimator that was initially proposed by T. Domencich and D. McFadden (1975) and studied by T. Amemiya (1978) in the context of the multivariate logit model. This procedure is also used for the estimation of the nested logit model (see D. McFadden (1981)). The asymptotic distribution of the two-step estimator is derived for the general model (1) - (5). This estimator is consistent but in general inefficient, and we characterize the cases for which the estimator is efficient. Then we consider two iterative sequential procedures. An iterative sequential procedure is a procedure in which the two-step or sequential estimator is applied at each iteration. The first iterative sequential procedure that we study basically relies on the adjustment, at each iteration, of the individual responses to A and B in such a way that the marginal frequencies of B for the adjusted responses agree with the marginal probabilities of B estimated at the previous iteration. It is shown, however, that the estimator obtained at any iteration has the same asymptotic distribution as the two-step estimator irrespective of which initial estimate is used to start the procedure. Another iterative sequential procedure which relies on a more complex adjustment of the individual responses to A and B is then proposed. The asymptotic distribution of the estimator obtained at each iteration is derived. It is shown that, under a certain condition, the estimator obtained at each iteration is more efficient that the estimator obtained at the previous iteration. Moreover, upon convergence of the procedure, the estimator hence obtained is shown to be asymptotically as efficient as the FIML estimator and therefore asymptotically efficient.

The paper is organized as follows. In Section 1, we introduce the notations and we consider the special case in which there are no relevant observed individual characteristics. In Sections 2, 3, and 4 we successively study, for that special case, the two-step estimator, a natural iterative sequential procedure, and an efficient iterative sequential procedure. In Section 5, these procedures and their properties are extended to the general model (1) - (5). Section 6 summarizes our results.

1. A Special Case

In this section, as well as in the next three sections, we consider the special case in which the (observed) individual characteristics are irrelevant. This means that the conditional probabilities $\Pr(A = i \mid B = j, t)$ and $\Pr(B = j \mid t)$ do not depend on the index t. The purpose of this simplifying assumption is to focus on the problems that are associated with the basic structure of the model.[4] Let us note that we may still have some characteristics of the choices.

Let Z_A be the $IJ \times a$ matrix of which the (i, j)-th row is z'_{ij}. Similarly, let Z_B be the $J \times b$ matrix of which the j-th row is z'_j. We shall use the notation e^n_N to indicate the n-th standard basis vector of R^N, and U_N to indicate the N-dimensional vector of ones. For any $j_o = 1,\ldots,J$, let ν^{j_o} be the IJ-dimensional vector of which the (i, j)-th component is one if $j = j_o$ and zero otherwise. Then $\nu^j = e^j_J \otimes U_I$ where " $\otimes$ " denotes the usual kronecker product.[5] Finally, let N_B be the $IJ \times J$ matrix of which the column vectors are ν^1 through ν^J.

Using the previous notation, the special case of the general model introduced in Section 1 can be written as:

$$\log \Pr(A \mid B) = \sum_{j=1}^{J} \mu_j \nu^j + Z_A \alpha \tag{6}$$

$$\log \Pr(B) = \mu U_J + Z_B \beta \tag{7}$$

where

$$Z_B = Z_B(\Pr(A|B)) \tag{8}$$

$$\mu_j = -\log\Big[\sum_{i=1}^{I} \exp z'_{ij}\alpha\Big] \tag{9}$$

$$\mu = -\log\Big[\sum_{j=1}^{J} \exp z'_j\beta\Big]. \tag{10}$$

[6]

We assume that the matrices M_A and M_B given by

$$M_A = [N_B, Z_A] = [\nu^1,\ldots,\nu^J, Z_A] \quad \text{and} \quad M_B = [U_J, Z_B]$$

are of full column rank, since if this were not the case, the parameters α and β would obviously not be identified. The linear manifolds generated by the column vectors of M_A and M_B are called the model spaces of the two probability models (6) and (7) (see footnote 6).

Let n_{ij} be the number of individuals whose A-response is equal to i, and B-response is equal to j. Since observed individual characteristics are irrelevant, the contingency table $n_{AB} = \{n_{ij};\ i = 1,\ldots,I,\ j = 1,\ldots,J\}$ is a sufficient statistic for the unknown parameter vector $\delta = (\alpha', \beta')'$.

We also assume, for the moment, that there are no observed empty cells in the contingency table, i.e., that $n_{ij} > 0$ for any pair (i, j),. Thus all the I.J conditional probabilities $p(i|j)$ and J marginal probabilities $p(j)$ can be estimated. Moreover, difficulties associated with the non-existence of M.L. estimates are removed.[7] The assumption of non-empty cells is all the more justified that we shall study the properties of estimators when the sample size becomes large.

The estimators we shall study are based on the maximum-likelihood method. We shall therefore consider the log-likelihood, more exactly, 1/T times the log-likelihood function. Let

$$L_A(\alpha,\ f_{AB}) = \sum_{i,j} f_{ij} \log p(i|j) \tag{11}$$

$$L_B(\alpha,\ \beta,\ f_B) = \sum_{j} f_j \log p(j) \tag{12}$$

where $f_{AB} = n_{AB}/T$ is the vector of observed frequencies, and $f_B = \{f_j;\ j = 1,\dots,J\}$ is the vector of marginal frequencies for B. The sample log-likelihood multiplied by 1/T is:

$$\begin{aligned} L(\alpha,\ \beta,\ f_{AB}) &= \sum_{i,j} f_{ij} \log p(i,\ j) \\ &= L_A(\alpha,\ f_{AB}) + L_B(\alpha,\ \beta,\ f_B) \end{aligned} \tag{13}$$

It is worth noting that $L(\alpha,\ \beta,\ f_{AB})$ is decomposed into the sum of two terms which are the log-likelihood functions one has to consider if the two probability models (3.1) and (3.2) are to be separately estimated by the maximum-likelihood method.

Three methods of estimation are now studied: a two-step (sequential) procedure, a natural iterative sequential procedure, and an efficient iterative sequential procedure. The proofs of the next

results as well as of complementary results are given in the Appendix. Expressions for the second order partial derivative of L_A and L_B are given in Lemma 3 of the Appendix. These expressions can be used to compute the asymptotic covariance matrices of the various estimators considered below.

2. A Two-Step Estimator

Given the structure of the model (6) - (7), it is natural to consider the two-step estimator $\hat{\delta}^1 = (\hat{\alpha}^{1\prime},\ \hat{\beta}^{1\prime})'$ of the true parameter vector $\delta^o = (\alpha^{o\prime},\ \beta^{o\prime})'$ that is described below. This two-step procedure was first proposed by T.A. Domencich and D. McFadden (1975) and studied by T. Amemiya (1978) in the context of a bivariate logit model. The same procedure is also used for the estimation of the nested logit model (see, e.g., D. McFadden (1981)). As shown in Q. H. Vuong (1982b), these models are special cases of the basic model (1)-(5).

In our case, the two-step procedure consists in:

(i) Estimating the probability model (6) for A given B by the method of maximum-likelihood (M.L.), i.e., maximizing $L(\alpha,\ f_{AB})$ with respect to α subject to the constraints (6) and (9). This gives an estimate $\hat{\alpha}^1$ and hence an estimate $\hat{Pr}^1(A|B)$.

(ii) Substituting $\hat{Pr}^1(A|B)$ for $Pr(A|B)$ in (8) to get an estimate $\hat{Z}^1_B$ of Z_B, and then estimating by the method of maximum-likelihood the probability model for B with $\hat{Z}^1_B$, instead of Z^1_B, i.e., maximizing $L(\hat{\alpha}^1, \beta, f_B)$ with respect to β subject to the constraints (7) and (10) where $\hat{Z}_B^1$ is substituted for Z_B^1. This gives the estimates $\hat{\beta}^1$ and $\hat{Pr}^1(B)$.

T. Amemiya (1978) has derived the asymptotic covariance matrix of the two-step estimator $\hat{\delta}^1 = (\hat{\alpha}^1, \hat{\beta}^1)$ in the case of a bivariate logit model. The next result shows that Amemiya's formula holds for the more general model (6) - (10). It is worth mentioning that the consistency and the asymptotic normality of the estimator $\hat{\alpha}^1$ result in fact from general properties of conditional maximum-likelihood estimators (E.B. Andersen (1970)). Indeed it is straightforward to see from the complete log-likelihood (13) that the marginal contingency table for B (or marginal frequencies f_B) is a sufficient statistic for the parameter vector β.

<u>Theorem 1</u>

The two-step estimator $\hat{\delta}^1$ is a consistent and asymptotically normal estimator of δ^o. And we have:

$$\sqrt{T}(\hat{\delta}^1 - \delta^o) \xrightarrow{D} N(0, \textstyle\sum^1)$$

where

$$\sum^1 = - \begin{bmatrix} \dfrac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + \dfrac{\partial^2 L_B}{\partial\alpha\partial\beta'} \left[\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \dfrac{\partial^2 L_B}{\partial\beta\partial\alpha'} & ; & \dfrac{\partial^2 L_B}{\partial\alpha\partial\beta'} \\ \\ \dfrac{\partial^2 L_B}{\partial\beta\partial\alpha'} & ; & \dfrac{\partial^2 L_B}{\partial\beta\partial\beta'} \end{bmatrix}^{-1}, \tag{14}$$

all the derivatives being evaluated at either $(\alpha^o, Pr^o(A, B))$ or $(\alpha^o, \beta^o, Pr^o(B))$.[8]

From the formula giving the inverse of a partitioned matrix, Equation (14) can be rewritten as:

$$\sum^1 = \begin{bmatrix} \sum^1_{\alpha\alpha} & \sum^1_{\alpha\beta} \\ \\ \sum^1_{\beta\alpha} & \sum^1_{\beta\beta} \end{bmatrix} \tag{15}$$

where

$$\sum^1_{\alpha\alpha} = - \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}$$

$$\sum^1_{\alpha\beta} = \sum^{1\,\prime}_{\beta\alpha} = \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} \left[\frac{\partial^2 L_B}{\partial\alpha\partial\beta'}\right] \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}$$

$$\sum^1_{\beta\beta} = - \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} - \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \left[\frac{\partial^2 L_B}{\partial\beta\partial\alpha'}\right] \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} \left[\frac{\partial^2 L_B}{\partial\alpha\partial\beta'}\right] \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}$$

As in Amemiya (1978), it is important to note that the asymptotic covariance matrix of the estimator $\hat{\alpha}^1$ is equal to the asymptotic covariance matrix of the estimator of α^o obtained from the separate M.L. estimation of the model (6).[9] However, due to the presence of the second term, the asymptotic covariance matrix of $\hat{\beta}^1$ is not equal to the asymptotic covariance matrix of the estimator of β^o obtained from the M.L. estimation of the model (7), assuming that $\alpha = \hat{\alpha}^1$. It is also noteworthy that $\left[\frac{-\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}$ is an underestimate of the true asymptotic covariance matrix $\sum^1_{\beta\beta}$ of β. This is because we have to take into account that $\hat{\alpha}^1$ is an estimate of α^o obtained in the first step of the procedure.

The two-step estimator $\hat{\delta}^1$ is in general inefficient. We shall, however, characterize the cases in which $\hat{\delta}^1$ is efficient. Let $\hat{\delta}^M = (\hat{\alpha}^{M\prime}, \hat{\beta}^{M\prime})'$ be the M.L. estimator of δ^o. It is well-known that under general conditions (which are satisfied here), the M.L. estimator satisfies:

$$\sqrt{T}(\hat{\delta}^M - \delta^o) \overset{D}{\rightarrow} N(0, \textstyle\sum^M)$$

where

$$\textstyle\sum^M = - \begin{bmatrix} \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} & ; & \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \\ \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} & ; & \frac{\partial^2 L_B}{\partial\beta\partial\beta'} \end{bmatrix}^{-1} , \tag{16}$$

all the matrices being evaluated at either $(\alpha^o, Pr^o(A, B))$ or $(\alpha^o, \beta^o, Pr^o(B))$.[10]

Let E be the square matrix defined by

$$E = \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} - \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} , \tag{17}$$

where all the derivatives are evaluated at $(\alpha, \beta, Pr(B ; \alpha, \beta))$. As lemma 6 of the Appendix states, the matrix E is negative semi-definite. Moreover $E = 0$ if and only if the column-vectors of $\frac{\partial Z_B\beta}{\partial\alpha'}$ belong to the linear manifold $M_B(\alpha)$ spanned by U_J and the column-vectors of $Z_B(\alpha)$. This condition can be written as:

$$\frac{\partial Z_B\beta}{\partial\alpha'} = M_B\,\Lambda \tag{18}$$

for some $(b + 1) \times a$ matrix Λ which may depend on α.

We can now readily evaluate the efficiency of the two-step estimator $\hat{\delta}^1$ by comparing its asymptotic covariance matrix to the asymptotic covariance matrix of the full-information maximum-likelihood (FIML) estimator $\hat{\delta}^M$.[11]

Corollary 1:

The two-step estimator $\hat{\delta}^1$ is in general inefficient, i.e.:

$$\sum^1 \geq \sum^M , \tag{19}$$

and is efficient if and only if (18) holds at (α^o, β^o).

Since the true parameters (α^o, β^o) are unknown, this latter result is used to check whether or not condition (18) holds for all parameters (α, β). As a matter of fact, if the dimension of the model space $M_B(\alpha)$ does not depend on α and thus is equal to b+1 so that all the parameters in β can be identified for any α , then it can readily be shown that condition (19) holds everywhere if and only if the model space $M_B(\alpha)$ is invariant with respect to α.

Corollary 2:

If the model space $M_B(\alpha)$ is invariant with respect to α, then the two-step estimator $\hat{\delta}^1$ is identical to the FIML estimator $\hat{\delta}^M$, and therefore efficient.

As discussed above, the sequential estimator $\hat{\delta}^1$ is consistent although in general inefficient. To obtain an efficient estimator, one can use a standard iterative optimization routine which will produce the FIML estimator $\hat{\delta}^M$. One can in fact carry out only one iteration of the Newton-Raphson algorithm (T.J. Rothenberg and T.C. Leenders (1964)) or one iteration of the algorithm of R. Berndt, B. Hall, R. Hall, J. Hausman (1974) on the complete log-likelihood function to obtain an efficient estimator of δ. Let us note that these procedures involve, however, the joint estimation of the parameter vector $\delta = (\alpha', \beta')'$.

In the following sections, we shall be interested in iterative sequential procedures, i.e., in procedures in which the parameter vectors α and β are sequentially estimated at each iteration. Such procedures will be particularly useful when, for instance, the number of parameters in α and β is large. Indeed, since the parameter vectors α and β are separately estimated in sequential procedures, the dimension of the problem will be reduced by approximately one half.

Given an estimate of α, one can readily obtain an estimate of β by applying the two-step procedure discussed earlier. The difficulty is to obtain a new estimate of α given an estimate of β, i.e., to use the estimate of β when considering the estimation of the model (6). One can distinguish three types of modifications to the estimation of model (6): one can modify (i) the observation vector n_{AB}, (ii) the vectors spanning Z_A, i.e., the values of the explanatory variables in model (6), and (iii) both n_{AB} and Z_A. We shall consider only the first type of modification.

3. A Natural Iterative Sequential Procedure

To define an iterative sequential procedure, we need to specify how the estimator $\hat{\delta}^{r+1} = (\hat{\alpha}^{r+1\prime}, \hat{\beta}^{r+1\prime})'$ of the r + 1 iteration is obtained from the estimator $\hat{\delta}^r = (\hat{\alpha}^{r\prime}, \hat{\beta}^{r\prime})$ of the previous iteration. Within any iteration, the estimator of β will be obtained from the estimator of α by applying the sequential method described in Section 2. To complete the definition of the sequential iterative procedure, it now suffices to describe how the estimator $\hat{\alpha}^{r+1}$ is obtained from $\hat{\delta}^r = (\hat{\alpha}^{r\prime}, \hat{\beta}^{r\prime})$. Let $\hat{Pr}^r(B)$ be the estimated marginal probability distribution of B that is associated with $\hat{\delta}^r$. Since the values of B are endogenously determined, and since an estimate of Pr(B) is available, it is natural to think of modifying the contingency table n_{AB} in such a way that the marginal frequencies of B agree with the estimate $\hat{Pr}^r(B)$.

Formally, if $\hat{f}^r_{AB}$ is the vector of adjusted frequencies, then

$$\hat{f}^r_{ij} = \frac{f_{ij}}{f_j} \cdot \hat{p}^r_j = f_{i|j} \cdot \hat{p}^r_j \tag{20}$$

or in matrix and vector notation:

$$\hat{f}^r_{AB} = D(\hat{Pr}^r(B) \otimes U_I) f_{A|B} = D(f_{A|B}) [\hat{Pr}^r(B) \otimes U_I]$$

where D(V) is the diagonal matrix of which the diagonal elements are the components of the vector V.[12] Let us note that this adjustment of the contingency table n_{AB} does not modify the conditional frequencies of A given B. On the other hand, the adjusted marginal frequency $\hat{f}^r_B$ is equal to $\hat{Pr}^r(B)$.

The estimator $\hat{\alpha}^{r+1}$ is then obtained by estimating the model for A given B where f_{AB} is replaced by the adjusted frequencies $\hat{f}^r_{AB}$.

The estimator $\hat{\alpha}^{r+1}$ will not in general be identical to the previous estimator $\hat{\alpha}^r$, even though the conditioual frequencies $f_{A|B}$ have not been modified. This is so because the conditional frequencies are not a sufficient statistic for α (see Equation (11)).

To summarize, the new estimator $(\hat{\alpha}^{r+1}, \hat{\beta}^{r+1})$ is obtained from $(\hat{\alpha}^r, \hat{\beta}^r)$ by:

(i) maximizing $L_A(\alpha, \hat{f}^r_{AB})$ with respect to α where $L_A(.,.)$ and $\hat{f}^r_{AB}$ are given by (11) and (20). This gives $\hat{\alpha}^{r+1}$.

(ii) maximizing $L_B(\hat{\alpha}^{r+1}, \beta, f_B)$ with respect to β, where $L_B(.,.,.)$ is given by (12). This gives $\hat{\beta}^{r+1}$.[13]

Let us note that these two steps require the estimation of two logit models so that standard programs can be used.

The iterative procedure is started by choosing an initial estimate of δ. As a matter of fact we only need an initial estimate of Pr(B). We shall restrict our attention to initial estimators $\hat{Pr}^o(B)$ that satisfy:

$$\sqrt{T}(\hat{Pr}^o(B) - Pr^o(B)) \xrightarrow{D} N(0,C) \tag{21}$$

for some covariance matrix C. One can for instance choose as an initial estimate of $\hat{Pr}^o(B)$, the observed marginal frequency f_B since it satisfied (21) (see lemma 1 of the Appendix). In this case, the first-iteration estimator $\hat{\delta}^1$ is simply the sequential estimator of Section 2.

The next theorem states, however, that the natural iterative sequential procedure does not lead to an efficient estimator of δ. More precisely, at any iteration this procedure provides an estimator of δ^o that is as efficient as the two-step estimator of Section 2.

<u>Theorem 2</u>

For any initial estimator of $Pr^o(B)$ that satisfies (21), and for any $r \geq 1$, the estimator $\hat{\delta}^r$ obtained at the r-th iteration of the natnral iterative sequential procedure has the same asymptotic distribution as the two-step estimator.

In particular, if we start from the initial consistent estimator f_B of $Pr^o(B)$, or from the two-step estimator $\hat{\delta}^1$ of Section 2, we get at each step of the procedure an estimator $\hat{\delta}^r$ which is asymptotically as efficient as the two-step estimator $\hat{\delta}^1$ which is in general inefficient. Let us note that the conclusion of Theorem 2 holds for any initial estimator that satisfied (21). Thus we may or may not improve the initial estimator $\hat{\delta}^o$ depending on whether $\hat{\delta}^o$ is less or more efficient than the two-step estimator $\hat{\delta}^1$. Let us also note that even in the case in which we improve the initial estimate, Theorem 2 implies that it is (asymptotically) worthless to do more than one iteration.

As another consequence of Theorem 2, we can derive the asymptotic properties of the limiting estimator $\hat{\delta}^L = (\hat{\alpha}^{L'}, \hat{\beta}^{L'})'$, if the procedure converges. If convergence occurs, then the limiting

estimator is a fixed point of the process and thns satisfies the limiting equations:

$$\left.\frac{\partial L_A}{\partial \alpha}\right|_{(\hat{\alpha}^L,\ \hat{f}^L_{AB})} = 0$$
$$\left.\frac{\partial L_B}{\partial \beta}\right|_{(\hat{\alpha}^L,\ \hat{\beta}^L,\ f_B)} = 0 \tag{22}$$

where

$$\hat{f}^L_{AB} = D(f_{A|B})[\hat{Pr}^L(B) \otimes U_I]. \tag{23}$$

Since for any iteration r, the estimator $\hat{\delta}^r$ is as efficient as the two-step estimator $\hat{\delta}^1$, it is not surprising that the limiting estimator $\hat{\delta}^L$ is also as efficient as $\hat{\delta}^1$. The next result is, however, slightly more general since it states that any consistent estimator that satisfied the limiting equations (22) with (23), is as efficient as the two-step estimator. Hence, the properties of the limiting estimator do not depend on the choice of the iterative procedure used to solve Equations (22)-(23).

<u>Corollary 3:</u>

Any consistent estimator $\hat{\delta}^L$ that satisfies (22)-(23) is asymptotically as efficient as the two-step estimator.

Before considering an efficient procedure in the next section, we now discuss why the natural iterative sequential procedure does not provide an estimator that is more and more efficient as the number of iteration increases.[14] As defined above, at the beginning of iteration $r + 1$, the observation vector n_{AB} is modified in such a way that the adjusted frequency $\hat{f}_B^r$ agrees with the marginal probability distribution $\hat{Pr}^r(B)$ estimated at iteration r. But any iterative sequential procedure that relies on adjustments that modify only the marginal frequencies f_B (and hence that preserve the conditional frequencies $f_{A|B}$) cannot provide in general a more efficient estimator than the estimator obtained at the previous iteration. To see this, let us assume that the probability model for A given B is saturated, i.e., $M_A = R^{IJ}$ (there are as many explanatory variables as independent conditional probabilities $p(i|j)$). It is easy to see from the first-order conditions (see Lemma 2 of the Appendix) that the estimated conditional probabilities $\hat{p}^r(i|j)$ are equal to the observed conditional frequencies $f_{i|j}$. Since the $f_{i|j}$'s are not modified, the estimate of α does not change, and hence neither does the estimate of β. Therefore in this case, an iterative sequential procedure based on the adjustment of only the marginal frequencies f_B gives at any step an estimator that is identical to the initial estimator.

4 An Efficient Iterative Sequential Procedure

The natural iterative sequential procedure previously discussed does not provide an estimator that is more efficient than the simple two-step estimator of Section 2. We shall now study another iterative sequential procedure in which the adjustments again

involve modifications only of the observation vector n_{AB}. As the discussion ending the preceding section suggests, we need to consider modifications that affect the conditional frequencies $f_{i|j}$ in order to obtain a useful iterative procedure.

The iterative sequential procedure studied in this section is similar to the previous natural iterative sequential procedure in the sense that it also relies on an adjustment of the observation vector in the first step of each iteration. However, the adjusted frequencies $\hat{f}_{AB}^r$ used in the first step of the r + 1 iteration is now:

$$\hat{f}_{AB}^r = f_{AB} - D(f_{AB})\, Z_A\left[\frac{\partial^2\hat{L}_A}{\partial\alpha\partial\alpha'}\right]^{-1} \frac{\partial L_B}{\partial\alpha}\bigg|_{(\hat{\alpha}^r,\ \hat{\beta}^r,\ f_B)} \tag{24}$$

where the matrix in square brackets is <u>any</u> consistent estimator of the second partial derivatives of L_A evaluated at $(\alpha^o, Pr^o(A, B))$. We shall use the estimate obtained by evaluating the second partial derivatives of L_A at $(\hat{\alpha}^1, f_{AB})$. Other consistent estimators are, of course, available, such as the second partial derivatives of L_A evaluated at $(\hat{\alpha}^r, f_{AB})$, or the estimate obtained by simply applying formulae (A.8) – (A.11) in the appendix where $Pr^o(A|B)$ and $Pr^o(B)$ are replaced by the consistent estimates $f_{A|B}$ and f_B. For practical reasons, however, the choice $(\hat{\alpha}^1, f_{AB})$ is preferable to $(\hat{\alpha}^r, f_{AB})$ since the $IJ \times a$ matrix by which the first order partial derivatives of L_B are premultiplied in (24) does not depend on $(\hat{\alpha}^r, \hat{\beta}^r)$ and hence need not be recomputed at each iteration.[15] Also, the method that consists in substituting $f_{A|B}$ and f_B for $Pr^o(A|B)$ and $Pr(B)$ in (A.8) – (A.11), although being very simple, may raise some difficulties when one does not have repeated observations (see footnote 27).

Since the first order partial derivatives of L_B are given by:

$$\left.\frac{\partial L_B}{\partial \alpha}\right|_{(\hat{\alpha}^r,\ \hat{\beta}^r,\ f_B)} = \left.\frac{\partial \beta' Z_B'}{\partial \alpha}\right|_{(\hat{\alpha}^r,\ \hat{\beta}^r)} [f_B - \hat{P}r^r(B)] \qquad (25)$$

(See Lemma 2 of the Appendix), it follows that the adjustment of the observed frequencies depends on the discrepancy between the observed marginal frequency f_B and the estimated probability distribution $\hat{P}r^r(B)$. However, since A depends on B so that Z_A is not of the form $U_J \otimes X_A$ for some $I \times a$ matrix X_A, then the adjustment given by (24) does modify the observed conditional frequency $f_{A|B}$.

To summarize, the iterative sequential procedure is as before, and thus consists in:

(i) maximizing $L_A(\alpha, \hat{f}^r_{AB})$ with respect to α, in order to obtain $\hat{\alpha}^{r+1}$.

(ii) maximizing $L_B(\hat{\alpha}^{r+1}, \beta, f_B)$ with respect to β in order to obtain $\hat{\beta}^{r+1}$.[16]

The only difference is that the adjusted frequency $\hat{f}^r_{AB}$ is now given by (24) instead of (20).

To start the iterative procedure, we shall use f_B as an initial (consistent) estimate of the marginal probability distribution Pr(B). From (24)-(25) it follows that the frequency vector $\hat{f}^o_{AB}$ used

at the first iteration is equal to the observed frequency f_{AB}. Thus the first iteration of the present procedure with f_B as an initial estimate simply gives the two-step estimator $\hat{\delta}^1$ of Section 2.

The next result basically gives the covariance matrix of the asymptotic distribution of the estimator obtained at each iteration of the iterative sequential procedure associated with the adjustment (24). Let F and G be the two square matrices defined as follows:

$$F = \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + E, \qquad (26)$$

$$G = E\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}, \qquad (27)$$

where E is given by (17), and where the partial derivatives of L_A are evaluated at $(\alpha, Pr(A, B; \alpha, \beta))$. Note that G is not necessarily symmetric while F is.

<u>Theorem 3:</u>

For any $r \geq 1$, the estimator $\hat{\delta}^r$ obtained at the r-th iteration of the iterative sequential procedure associated with the adjustment (24) is a consistent and asymptotically normal estimator of δ^o. And we have:

$$\sqrt{T}(\hat{\delta}^r - \delta^o) \xrightarrow{D} N(0, \textstyle\sum^r)$$

where

$$\sum^{r} = - \begin{bmatrix} (I + G^{2r-1})^{-1} F + \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} (\frac{\partial^2 L_B}{\partial\beta\partial\beta'})^{-1} \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} ; & \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \\ \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} ; & \frac{\partial^2 L_B}{\partial\beta\partial\beta'} \end{bmatrix}^{-1} . \quad (28)$$

The matrices H and F, and all the partial derivatives are evaluated at either (α^o, $Pr^o(A, B)$) or (α^o, β^o, $Pr^o(B)$).[17]

Let us note that

$$F = (I + G) \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} , \quad (29)$$

Thus if $r = 1$, then $\sum^{r}$ is equal to the covariance matrix (3.9), as it should be since $\hat{\delta}^1$ is simply the two-step estimator.

Equation (28) shows that the asymptotic covariance matrix of $\hat{\delta}^r$ depends in general on r. As mentioned above, the matrix G is not necessarily symmetric. However, it is shown in the Appendix, Lemma 14, that all the roots of G are real and nonnegative. The next corollary states that one improves the estimator of δ^o at each iteration if and only if the largest root of G^o [G^o is the matrix G evaluated at (α^o, β^o, $Pr^o(A, B)$)] is strictly less than one.[18] Specifically, the estimator obtained at each iteration is more efficient than the estimator obtained at the previous iteration.

Corollary 4 also compares the efficiency of each estimator to the FIML estimator which is efficient.

<u>Corollary 4:</u>

For any r, the estimator $\hat{\delta}^{r+1}$ is at least as efficient as the estimator $\hat{\delta}^r$ if and only if the largest roots of G^o is not greater than one, and is strictly more efficient than $\hat{\delta}^r$ if and only if the largest root of G^o is strictly less than one.

Moreover, the FIML estimator $\hat{\delta}^M$ is at least as efficient as $\hat{\delta}^r$ for any r.

Hence, when the largest root of G^o is strictly less than one, we have:

$$\sum^{M} > \sum^{r+1} > \sum^{r} \quad (30)$$

Thus, unlike the natural iterative sequential procedure, it pays to iterate with the present procedure. In fact, the present iterative sequential procedure is efficient in the sense that the asymptotic covariance matrix of $\hat{\delta}^r$ converges to the inverse of the information matrix. More precisely, this holds under the condition of Corollary 4.

Corollary 5:

When r increases, $\hat{\delta}^r$ becomes efficient, i.e.:

$$\lim_{r\to\infty} \sum^r = \sum^M \tag{31}$$

if and only if the largest root of G^o is strictly less than one.

Since the true parameters α^o and β^o, and hence the matrix G^o are not known, the conditions of Corollaries 4 and 5 cannot be checked. Of course, one may impose that for all (α, β), the matrix G has its largest root less than one. But this latter condition may be violated by too many models. However, a practical way to know if the largest root of G^o is smaller than one is to look at the convergence of the iterative sequential procedure. Indeed if the largest root of G^o is strictly less than one, and if the initial estimator is sufficiently close to the true value of the parameters, then convergence will occur. In this case, the limiting estimator $\hat{\delta}^L$ is efficient and satisfied the limiting equations (22) with

$$\hat{f}^L_{AB} = f_{AB} - D(f_{AB})\, Z_A \left[\frac{\partial^2 \hat{L}_A}{\partial\alpha\partial\alpha'}\right]^{-1} \frac{\partial L_B}{\partial\alpha}\bigg|_{(\hat{\alpha}^L,\ \hat{\beta}^L,\ f_B)} \tag{32}$$

The next corollary states that any consistent estimator of δ that is a solution of the limiting equations (22) with $\hat{f}^L_{AB}$ as defined by (32) is an efficient estimator of δ^o. Thus the efficiency property

holds irrespective of which iterative procedure is used to solve the limiting equations.[19]

Corollary 6:

Any consistent estimator $\hat{\delta}^L$ that satisfies (22) and (32) is efficient.

5. Generalization

In the previous sections, we have studied various estimation procedures for the case in which there are no observed relevant individual characteristics so that the contingency table n_{AB} is a sufficient statistic. In addition, we have assumed that there are no observed empty cells in the contingency table in order to study the asymptotic properties of the proposed estimators. For a fixed sample size, however, this latter condition is often not satisfied.

The purpose of this section is to see whether the properties of the previous estimation procedures, and in particular the efficient iterative sequential procedure, still hold when the contingency table has some empty cells or when there are some observed relevant individual characteristics. First, under the assumption that there are no relevant individual characteristics, we consider the case in which the contingency table has some empty cells. We then extend the discussion to the case in which there are some relevant observed individual characteristics that are all qualitative. Finally, we discuss the case in which some individual characteristics are quantitative.

When there are some observed empty cells, two major difficulties arise: a problem of identification of the parameters α, and a problem of existence of M.L. estimates. Let us for instance consider the two-step estimator $\hat{\delta}^1$ for which the estimate $\hat{\alpha}^1$ is obtained by maximizing $L_A(\alpha, f_{AB})$ as given by (11). Suppose that some categories of B are not observed. Then from (11) it follows that the corresponding conditional probabilities $p(i|j)$ do not appear in the (conditional) likelihood function L_A. If the conditional probability model for A given B is too large, then there may exist more than one (actually an infinite number of) admissible conditional probability distributions for A given B that disagree only on the unobserved values of B. If this is the case, then clearly one cannot have a unique estimate $\hat{\alpha}^1$. It can, however, be readily shown that this does not occur (the conditional likelihood L_A being then strictly concave in α) if and only if the matrix M_A^* is full column-rank, where M_A^* is the matrix obtained from M_A by deleting all the rows and columns associated with the unobserved categories of B.[20] In what follows, it is assumed that this condition holds.

The second problem is well known and is associated with the existence of the solutions of the two maximization problems that have to be solved at each step of the iterative sequential procedure. Indeed, since some f_{ij} and some f_j are equal to zero, the estimates $\hat{\alpha}^r$ and $\hat{\beta}^r$ may not exist.[21] If this is the case, then the only solution

is to obtain a larger sample, or to reduce the set of admissible probability distributions, i.e., the probability model, by introducing some additional restrictions on the parameters.

We now introduce some relevant individual characteristics. We first assume that these individual characteristics are all qualitative. Without loss of generality, we can consider only one qualitative characteristic C that is polytomous. Let K be the number of its categories.[22] The recursive system of probability models with probability feedback becomes

$$\log \Pr(A|B, C = k) = \sum_j \mu_{jk} V^j + Z_{Ak}\alpha \tag{33}$$

$$\log \Pr(B|C = k) = \mu_k U_J + Z_{Bk}\beta \tag{34}$$

where Z_{Bk} depends on α.

Let f_{ABC} be the vector of observed frequencies for A, B, and C. Also, let $f_{AB|C}$ and f_C be the vector of conditional frequencies for A and B given C, and the vector of marginal frequencies for C. Then, the two log-likelihood functions that are considered in the two-step estimator or in the iterative sequential procedures previously studied are:

$$L_A(\alpha,\ f_{AB|C},\ f_C) = \sum_k f_k\, L_{Ak}(\alpha,\ f_{AB|k}) \tag{35}$$

$$L_B(\alpha,\ \beta,\ f_{B|C},\ f_C) = \sum_k f_k\, L_{Bk}(\alpha,\ \beta,\ f_{B|k}) \tag{36}$$

where

$$L_{Ak}(\alpha,\ f_{AB|k}) = \sum_{i,j} f_{ij|k} \log p(i|j,k) \tag{37}$$

$$L_{Bk}(\alpha,\ \beta,\ f_{B|k}) = \sum_{j} f_{j|k} \log p(j|k) \tag{38}$$

The summations in (35) and (36) should of course be taken over the values of C that are observed. If some values of C are not observed or if some combinations of B and C are not observed, as will frequently be the case when the number of categories of C is large relative to the sample size, then problems of identification of some parameters in α and β may arise.[23] Moreover, problems associated with the existence of values for α and β that maximize $L_A(\alpha,\ f_{ABC})$ and $L_B(\alpha,\ \beta,\ f_{BC})$ may occur. Besides these two difficulties which were discussed above, we have to reconsider the adjustments that are made in the efficient iterative sequential procedure.[24] We shall apply equation (24) in order to adjust each conditional frequency $f_{AB|k}$. Specifically, for any value k of C that is observed, we have:

$$\hat{f}^r_{AB|k} = f_{AB|k} - D(f_{AB|k})\ Z_{Ak}\left[\frac{\partial^2 \hat{L}_{Ak}}{\partial\alpha\partial\alpha'}\right]^{-1} \frac{\partial L_{Bk}}{\partial\alpha}\Bigg|_{(\hat{\alpha}^r,\ \hat{\beta}^r,\ f_{B|k})} \tag{39}$$

where the matrix in square brackets is a consistent estimator of the second partial derivatives of L_{Ak} at $(\alpha^o,\ Pr^o(A,\ B|k))$. We shall take as consistent estimator the second partial derivatives of L_{Ak} evaluated at $(\hat{\alpha}^1,\ f_{AB|k})$ where $\hat{\alpha}^1$ is the two-step estimator of α. Let

us note that the above adjustment modifies only the conditional frequencies $f_{AB|C}$. Hence $\hat{f}^r_C = f_C$.

The first partial derivatives of L_{Bk} with respect to α are:

$$\frac{\partial L_{Bk}}{\partial\alpha}\Bigg|_{(\hat{\alpha}^r,\ \hat{\beta}^r,\ f_{B|C})} = \frac{\partial\beta' Z'_{Bk}}{\partial\alpha}\Bigg|_{(\hat{\alpha}^r,\ \hat{\beta}^r)} (f_{B|k} - \hat{Pr}^r(B|k)). \tag{40}$$

Hence, if the observed conditional frequency $f_{B|C}$ is taken to be the initial estimate of $Pr(B|k)$, then $\hat{f}^1_{ABC} = f_{ABC}$ so that the estimator $\hat{\delta}^1$ obtained at the first iteration is again the two-step estimator.

For the study of the asymptotic properties, it is assumed that the observed marginal frequency f_C converges in probability to some probability distribution $Pr^o(C)$.[25] Then it can readily be shown (see Appendix, Section 5) that all the results of Sections 2, 3, and 4 still hold.[26] In particular the iterative sequential procedure defined by the adjustment (39) used in the first step of each iteration is efficient in the sense that it produces an asymptotically efficient estimator of α and β when the number of iterations increases. Moreover, the asymptotic covariance matrix of the estimator obtained at the r-th iteration is given by (28) where all the second partial derivatives are evaluated at $(\alpha^o,\ \beta^o,\ Pr^o(A,B|C),\ Pr^o(C))$. These partial derivatives can readily be evaluated from (35) - (36) by noting that Lemma 3 of the Appendix applies to L_{Ak} and L_{Bk} for each k. For instance, we have:

$$\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\bigg|_{(\alpha,\ \beta,\ Pr^o(A,B|C),\ Pr^o(C))} = \sum_k p_k^o \frac{\partial^2 L_{Ak}}{\partial\alpha\partial\alpha'}\bigg|_{(\alpha,\ \beta,\ Pr^o(A,B|k))} \tag{41}$$

where the second partial derivatives of L_{Ak} are given by Equation (A.11).[27]

The study of the case in which all the individual characteristics are qualitative clearly suggests how one can proceed when some or all of these characteristics are continuous. In the latter case, the recursive system with probability feedback takes the general form of Equations (1) - (5). Let y_{ABt} be the IJ-dimensional vector of which the (i,j)-th component is:

$$y_{ijt} = \begin{cases} 1 & \text{if } A_t = i \quad \text{and} \quad B_t = j \\ 0 & \text{otherwise} \end{cases} \tag{42}$$

Let y_{Bt} be the J-dimensional vector of which the j-th element y_{jt} is the sum of the y_{ijt}'s over i. Then

$$y_{jt} = \begin{cases} 1 & \text{if } B_t = j \\ 0 & \text{otherwise} \end{cases} \tag{43}$$

Define y_{ABX} and y_{BX} as being the vectors $(y'_{AB1},\ldots,y'_{ABT})'$ and $(y'_{B1},\ldots,y'_{BT})'$. Then the two likelihood functions that we consider are:

$$L_A(\alpha,\ y_{ABX}) = \frac{1}{T}\sum_t L_{At}(\alpha,\ y_{ABt}) \tag{44}$$

$$L_B(\alpha,\ \beta,\ y_{BX}) = \frac{1}{T}\sum_t L_{Bt}(\alpha,\ \beta,\ y_{Bt}) \tag{45}$$

where

$$L_{At}(\alpha,\ y_{ABt}) = \sum_{i,j} y_{ijt} \log p(i|j,t) \tag{46}$$

$$L_{Bt}(\alpha,\ \beta,\ y_{Bt}) = \sum_j y_{jt} \log p(j|t) \tag{47}$$

Comparing Equations (44) - (47) to Equations (35) - (38) it is clear that t plays the role of k while y_{ABt} and y_{Bt} play the role of $f_{AB|k}$ and $f_{B|k}$. Hence, provided the continuous characteristics satisfy some suitable asymptotic assumptions (such as those given by, e.g., D. McFadden (1974) or C. Manski and D. McFadden (1981)), one can show that all the results of Sections 2 - 4 still hold. We shall not, however, elaborate on these technicalities.[28] Instead, we now briefly summarize the results of Sections 2 - 4 in this more general context.

In the two-step procedure, we first obtain an estimate $\hat{\alpha}^1$ by maximizing (44) with respect to α, then we obtain an estimate of β by maximizing $L_B(\hat{\alpha}^1,\ \beta,\ y_{BX})$ with respect to β. According to the result of Section 2, this two-step estimator is consistent, asymptotically normal, but in general inefficient.[29] Moreover, the only practical cases for which the two-step estimator is efficient (and equal to the FIML estimator) are those for which the linear space M_B does not depend on α, where M_B is generated by the JT-dimensional vectors

$\{e_T^t \otimes U_J\ ;\ \ t = 1,\ldots,T\}$, and the b column-vectors of the matrix $Z_B = [Z'_{B1},\ldots, Z'_{BT}]'$ (see Appendix, Section 5).

In the natural iterative sequential procedure, an adjustment of y_{ABX} is made in the first step of any iteration. This adjustment is defined for any observation t and for any iteration r as follows:

$$\hat{y}^r_{ijt} = \begin{cases} \hat{p}_{jt} & \text{if } A_t = i \quad \text{and} \quad B_t = j \\ 0 & \text{otherwise} \end{cases} \tag{48}$$

(Compare (48) to (43).) In the r + 1 iteration, $\hat{\alpha}^{r+1}$ is obtained by maximizing $L_A(\alpha, \hat{y}^r_{ABX})$ with respect to α, then $\hat{\beta}^{r+1}$ is obtained by maximizing $L_B(\hat{\alpha}^{r+1}, \beta, y_{BX})$ with respect to β. Theorem 2 says that the estimator obtained at each iteration is as efficient as the two-step estimator irrespective of which initial estimates are used to start the procedure.

Finally, let us consider the efficient iterative sequential procedure. It follows from (39) and (42), that the adjustment is:

$$\hat{y}^r_{ijt} = \begin{cases} 1 - z'_{ijt}\left[\dfrac{\partial^2 L_{At}}{\partial\alpha\partial\alpha'}\right]^{-1} \dfrac{\partial L_{Bt}}{\partial\alpha}\Bigg|_{(\hat{\alpha}^r,\ \hat{\beta}^r,\ y_{Bt})} \\ \qquad\qquad \text{if } A_t = i \quad \text{and} \quad B_t = j \\ 0 \quad \text{otherwise} \end{cases} \tag{49}$$

where the matrix in square brackets is a consistent estimate of the second partial derivatives of L_{At} at $(\alpha^o, Pr^o(A,B|t))$. We shall use

as a consistent estimate the second partial derivatives of L_{At} at $(\hat{\alpha}^1, y_{ABt})$. We have

$$\frac{\partial^2 L_{At}}{\partial\alpha\partial\alpha'}\Bigg|_{(\hat{\alpha}^1,\ y_{ABt})} = -Z'_{At}\, D(y_{Bt} \otimes U_I)\, \hat{\Omega}^1_{A|Bt}\, Z_{At} \tag{50}$$

where $\hat{\Omega}^1_{A|Bt}$ is given by formula (A.10) of the Appendix with $\hat{Pr}^1(A|B,t)$ substituted for $Pr(A|B)$.[30] The IJ $\times$ a matrix Z_{At} can be partitioned into J submatrices Z_{Ajt}, j = I,...,J according to the values of B. From (43), (49), and (50), it follows that the adjustment used in the efficient iterative sequential procedure is:

$$\hat{y}^r_{ijt} = \begin{cases} 1 + z'_{ijt}\left[Z'_{Ajt}\, \hat{\Omega}^1_{A|jt}\, Z_{Ajt}\right]^{-1} \dfrac{\partial\beta' Z'_{Bt}}{\partial\alpha}\Bigg|_{(\hat{\alpha}^r,\ \hat{\beta}^r)} (y_{Bt} - \hat{Pr}^r(B|t)) \\ \qquad\qquad \text{if } A_t = i \quad \text{and} \quad B_t = j \\ 0 \quad \text{otherwise} \end{cases} \tag{51}$$

where $\hat{\Omega}^1_{A|jt}$ is given by formula (A.9) of the Appendix with $\hat{Pr}^1(A|B,t)$ substituted for $Pr(A|B)$.

Theorem 3 then implies that the estimator obtained at each iteration of this procedure is consistent and asymptotically normal. Its asymptotic covariance matrix is given by (28).[31] Furthermore, when the number of iterations increases, the procedure produces, upon convergence, an asymptotically efficient estimator.

6. Conclusion

In this paper we have studied some methods for estimating a recursive system of logit models with probability feedback. The asymptotic distribution of the two-step estimator was derived for this general model, and the cases for which the estimator is efficient were characterized. Two iterative sequential procedures were also considered. The first one relies on an adjustment, at each iteration, of the individual responses in such a way that they agree with the marginal probabilities estimated at the previous iteration. Unfortunately, this simple procedure produces, at every iteration, an estimator that is only as efficient as the two step-estimator. A more complex adjustment of the individual responses was then proposed. The asymptotic distribution of the estimator obtained at each iteration of the associated iterative sequential procedure was derived. Moreover, it was shown that, upon convergence, the procedure produces an asymptotically efficient estimator.

The two iterative sequential procedures that we studied both modify the individual responses. We have mentioned that other iterative sequential procedures can be constructed by adjusting the individual responses or the explanatory variables. An appropriate goal for further research would be to find an iterative sequential procedure that would generate an efficient estimator in two iterations. Also, our results were derived for the case in which the probability models of the resursive system belong to the class of logit models. It is likely, however, that the results continue to hold when the component models belong to a broader class of probability models.

APPENDIX

1. Some Preliminary Results

The large sample properties given in Sections 2 - 4 are based on the following well-known asymptotic properties of the observed frequencies f_{AB} when the observations are mutually independent, i.e., when we have a random sample (e.g., C. R. Rao (1965)). Let Ω be the IJ square matrix defined by:

$$\Omega = D(Pr(A, B)) - Pr(A, B) \,.\, Pr(A, B)' \tag{A.1}$$

where D(V) is the diagonal matrix of which the diagonal elements are the components of the vector V. Denote by Ω^o the value of Ω at the true joint probability distribution $Pr^o(A, B)$.

Lemma 1:

$$\text{plim } f_{AB} = Pr^o(A, B) \tag{A.2}$$

$$\sqrt{T}\left[f_{AB} - Pr^o(A, B)\right] \xrightarrow{D} N(0, \Omega^o). \tag{A.3}$$

Let us note that the asymptotic properties of the marginal frequencies f_B and the conditional frequencies $f_{A|B}$ can readily be derived from this lemma.

The next lemma gives the expressions of the first partial derivatives of $L_A(\alpha, f_{AB})$ and $L_B(\alpha, \beta, f_B)$ evaluated at (α, f_{AB}) and

(α, β, f_B) respectively.

Lemma 2:

The first partial derivatives of $L_A(\alpha, f_{AB})$ and $L_B(\alpha, \beta, f_B)$ evaluated at (α, f_B) and (α, β, f_B) respectively are:

$$\frac{\partial L_A}{\partial \alpha} = Z_A'\left[f_{AB} - D(f_B \otimes U_I)Pr(A|B)\right] \tag{A.4}$$

$$\frac{\partial L_B}{\partial \alpha} = \frac{\partial \beta' Z_B'}{\partial \alpha}\left[f_B - Pr(B)\right] \tag{A.5}$$

$$\frac{\partial L_B}{\partial \beta} = Z_B'\left[f_B - Pr(B)\right] \tag{A.6}$$

proof:

Equations (A.4) and (A.6) are well-known (see, e.g., S. J. Haberman (1974)). To establish (A.5), we compute the first partial derivative of log Pr(B) with respect to α. From (3.1), we get:

$$\frac{\partial \log Pr(B)}{\partial \alpha'} = U_J \frac{\partial \mu}{\partial \alpha'} + \frac{\partial Z_B \beta}{\partial \alpha'},$$

and from (3.5):

$$\frac{\partial \mu}{\partial \alpha'} = - Pr(B)' \frac{\partial Z_B \beta}{\partial \alpha'}$$

Thus

$$\frac{\partial \log Pr(B)}{\partial \alpha'} = \left[I - U_J \,.\, Pr(B)'\right] \frac{\partial Z_B \beta}{\partial \alpha'}$$

Since $L_B = f_B' \,.\, \log Pr(B)$, we have the desired result.

Q. E. D.

We can now readily compute the second partial derivatives of $L_A(\alpha, f_{AB})$ and $L_B(\alpha, \beta, f_{AB})$. Let

$$\Omega_B = D(Pr(B)) - Pr(B) \,.\, Pr(B)' \tag{A.7}$$

$$V_{A|B} = D(Pr(B) \otimes U_I)\, \Omega_{A|B} \tag{A.8}$$

The IJ square matrix $\Omega_{A|B}$ is block-diagonal with the j-th block equal to the I-square matrix:

$$\Omega_{A|j} = D(Pr(A|B = j)) - Pr(A|B = j) \,.\, Pr(A|B = j)' \tag{A.9}$$

Or equivalently, we have:

$$\Omega_{A|B} = D(Pr(A|B)) - D(Pr(A|B))\, N_B N_B'\, D(Pr(A|B)) \tag{A.10}$$

where N_B is the IJ $\times$ I matrix defined at the beginning of Section 2.

Lemma 3:

(i) The second partial derivatives of $L_A(\alpha, f_{AB})$ evaluated at $(\alpha, Pr(A, B))$ are:

$$\frac{\partial^2 L_A}{\partial \alpha \partial \alpha'} = - Z_A' V_{A|B} Z_A \tag{A.11}$$

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} = Z'_A\left[I - D(Pr(A|B))(N'_B \otimes U_I')\right] \tag{A.12}$$

(ii) The second partial derivatives of $L_B(\alpha,\ \beta,\ f_B)$ evaluated at $(\alpha,\ \beta,\ Pr(B))$ are:

$$\frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} = -\frac{\partial\beta' Z'_B}{\partial\alpha}\,\Omega_B\,\frac{\partial Z_B\beta}{\partial\alpha'} \tag{A.13}$$

$$\frac{\partial^2 L_B}{\partial\alpha\partial\beta'} = -\frac{\partial\beta' Z'_B}{\partial\alpha}\,\Omega_B\,Z_B \tag{A.14}$$

$$\frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} = \frac{\partial\beta' Z'_B}{\partial\alpha}\,N'_B \tag{A.15}$$

$$\frac{\partial^2 L_B}{\partial\beta\partial\beta'} = -Z'_B\,\Omega_B\,Z_B \tag{A.16}$$

$$\frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}} = Z'_B\,N'_B \tag{A.17}$$

where I is the identity matrix of which the order can be easily determined from the context.

proof:

Equation (A.16) is a standard result (see, e.g., S. J. Haberman (1974)). Equations (A.12), (A.15), and (A.17) can be straightforwardly derived from (A.4), (A.5), and (A.6) by noting that:

$$\frac{\partial f_B}{\partial f'_{AB}} = N'_B\ .$$

To establish (A.11), we differentiate (A.4) with respect to α and we use

$$\frac{\partial\ Pr(A|B)}{\partial\alpha'} = \Omega_{A|B}\,Z_A\ ,$$

to obtain:

$$\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} = -Z'_A\,D(f_B \otimes U_I)\,\Omega_{A|B}\,Z_A\ .$$

Finally, since

$$\frac{\partial\ Pr(B)}{\partial\alpha'} = \Omega_B\,\frac{\partial Z_B\beta}{\partial\alpha'}\ ;\quad \frac{\partial\ Pr(B)}{\partial\beta'} = \Omega_B\,Z_B$$

(see the proof of Lemma 2), then by differentiating (A.5) with respect to either α or β we get:

$$\frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} = -\frac{\partial\beta' Z'_B}{\partial\alpha}\,\Omega_B\,\frac{\partial Z_B\beta}{\partial\alpha'} + \frac{\partial}{\partial\beta'}\left[\frac{\partial\beta' Z'_B}{\partial\alpha}\right] \cdot [f_B - Pr(B)]$$

$$\frac{\partial^2 L_B}{\partial\alpha\partial\beta'} = -\frac{\partial\beta' Z'_B}{\partial\alpha}\,\Omega_B\,Z_B + \frac{\partial}{\partial\beta'}\left[\frac{\partial\beta' Z'_B}{\partial\alpha}\right] \cdot [f_B - Pr(B)]$$

Thus at $(\alpha,\beta,Pr(B))$, the second term of each equation vanishes. This establishes (A.13) and (A.14).

Q. E. D.

As a matter of fact, Equations (A.12), (A.15), (A.16), and (A.17) also hold when the partial derivatives are evaluated at (α, f_{AB}) or (α, β, f_B). However, as the proof of Lemma 3 shows, Equations (A.11), (A.13), and (A.14) hold only at $(\alpha, \beta, Pr(A, B))$. It is worth noting that the second partial derivatives (A.13) and (A.14) do not involve the second partial derivatives of Z_B. Let us also note that the partial derivatives of L_B are taken with respect to f_{AB}, not f_B.

The properties stated in the next lemma are used to compute products of partial derivatives.

<u>Lemma 4:</u>

$$\Omega_B = N_B' \, \Omega \, N_B \tag{A.18}$$

$$0 = [I - D(Pr(A|B))(N_B' \otimes U_I)] \, \Omega \, N_B \tag{A.19}$$

$$V_{A|B} = [I - D(Pr(A|B))(N_B' \otimes U_I)] \, \Omega \, [I - D(Pr(A|B))(N_B' \otimes U_I)]' \tag{A.20}$$

proof:

Equation (A.18) is straightforward from (A.1) and (A.7) since:

$$\begin{aligned} D(Pr(B)) &= N_B' \, D(Pr(A, B)) \, N_B \\ Pr(B) &= N_B' \, Pr(A, B) \end{aligned}$$

To establish (A.19) and (A.20), we shall use the fact that the kernel of $\Omega_{A|B}$ is the linear manifold spanned by the column vectors of N_B, i.e., the set of vectors of the form $V \otimes U_I$. Let us note that:

$$\begin{aligned} \Omega \, N_B &= D(Pr(A, B)) \, N_B - Pr(A, B) \, . \, Pr(A, B)' N_B \\ &= D(Pr(A|B))[D(Pr(B) \otimes U_I)N_B - (Pr(B) \otimes U_I)Pr(B)'] \\ &= D(Pr(A|B))[D(Pr(B) \otimes U_I - [Pr(B)Pr(B)'] \otimes U_I] \\ &= D(Pr(A|B))[\Omega_B \otimes U_I] \end{aligned}$$

Since $N_B' \otimes U_I = N_B N_B'$, it follows from (A.10) that the right-hand side of (A.19) is $\Omega_{A|B}[\Omega_B \otimes U_I]$. It now suffices to use the fact that the vectors of the form $V \otimes U_I$ belong to the kernel of $\Omega_{A|B}$ in order to establish (A.19).

Since $N_B' \otimes U_I = N_B N_B'$, then the transpose of the second square bracket in (A.20) is $I - N_B N_B' \, D(Pr(A|B))$. It follows from (A.19) that we must now show that $V_{A|B}$ is equal to the first square bracket of (A.20) post multiplied by Ω. This product is equal to:

$$[I - D(Pr(A|B))N_B'N_B] \, D(Pr(A|B)) \, [D(Pr(B) \otimes U_I) - (Pr(B) \otimes U_I)Pr(A, B)']$$

which is also equal to:

$$\Omega_{A|B} \, [D(Pr(B) \otimes U_I) - (Pr(B) \otimes U_I)Pr(A, B)']$$

because of (A.10). But $\Omega_{A|B}$ post multiplied by the second term in

f_{AB}. Moreover f_{AB} is a consistent estimator of $Pr^o(A, B)$ (Lemma 1), and (α^o, $Pr^o(A, B)$) satisfied the first equation (see (A.4)). Thus $\hat{\alpha}^1$ is a consistent estimator of α^o. Similarly, since L_B has continuous second partial derivatives, then $\hat{\beta}^1$ is a continuous function of $\hat{\alpha}^1$ and f_B. These latter estimators of α^o and $Pr^o(B)$ are consistent. Moreover (α^o, β^o, $Pr^o(B)$) satisfies the second normal equation (see (A.5)). Thus $\hat{\beta}^1$ is a consistent estimator of β^o.

The previous argument in fact also shows that both $\hat{\alpha}^1 - \alpha^o$ and $\hat{\beta}^1 - \beta^o$ are $O(T^{-1/2})$ because $f_{AB} - Pr^o(A, B)$ is $O(T^{-1/2})$ from Lemma 1 (for more information on the use of O and o which indicate the order in probability of a sequence of random variables, see H. B. Mann and A. Wald (1943)). We can now apply a Taylor expansion around the value (α^o, β^o, $Pr^o(A, B)$). Using the fact that this value also satisfies the normal equations, we get:

$$0 = \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}(\hat{\alpha}^1 - \alpha^o) + \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}(f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.28}$$

$$\begin{aligned} 0 = {} & \frac{\partial^2 L_B}{\partial\alpha\partial\beta'}(\hat{\alpha}^1 - \alpha^o) + \frac{\partial^2 L_B}{\partial\beta\partial\beta'}(\hat{\beta}^1 - \beta^o) \\ & + \frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}}(f_{AB} - Pr^o(A, B)) + O(T^{-1}) \end{aligned} \tag{A.29}$$

where the partial derivatives of L_A and L_B are evaluated at (α^o, $Pr^o(A, B)$) and (α^o, β^o, $Pr^o(B)$) respectively. Solving for ($\hat{\alpha}^1 - \alpha^o$) and ($\hat{\beta}^1 - \beta^o$), we get:

$$\hat{\alpha}^1 - \alpha^o = -\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}(f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.30}$$

$$\hat{\beta}^1 - \beta^o = -\left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \left[\frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}} - \frac{\partial^2 L_B}{\partial\beta\partial\alpha'}\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}\right] \tag{A.31}$$

Since $\sqrt{T}(f_{AB} - Pr^o(A, B)$ converges in distribution to $N(0, \Omega^o)$ (Lemma 1), it follows that $\sqrt{T}(\hat{\delta}^1 - \delta^o)$ converges in distribution to s normal distribution. To determine the asymptotic covariance matrix of $\hat{\delta}^1$, let us first derive the asymptotic covariance matrix of $\hat{\alpha}^1$. It follows from (A.2), (A.21), and (A.30) that $\sqrt{T}(\hat{\alpha}^1 - \alpha^o)$ converges in distribution to $N(0, \sum^1_{\alpha\alpha})$ where $\sum^1_{\alpha\alpha}$ is given below (15).

Similarly, from (A.2), (A.23), (A.26), and (A.31), it follows that $\sqrt{T}(\hat{\beta}^1 - \beta^o)$ converges in distribution to $N(0, \sum^1_{\beta\beta})$ where $\sum^1_{\beta\beta}$ is given below (15).

Finally to show that the asymptotic covariance matrix between $\hat{\alpha}^1$ and $\hat{\beta}^1$ is equal to the matrix $\sum^1_{\alpha\beta}$ given below (15), it suffices to use (A.30) - (A.31) with the equations (A.21), (A.23), and (A.26).

2.2 A Lemma

Corollary 1 uses the following properties of the square matrix E defined by (17).

Lemma 6:

The matrix E is negative semi-definite. Moreover, E = 0 if and only if

$$\frac{\partial Z_B\beta}{\partial\alpha'} = M_B\,\Lambda \qquad (A.32)$$

for some (b + 1) $\times$ a matrix Λ which may depend on α.

Proof:

To prove the negative semi-definiteness of E, we can use the positive semi-definiteness of the (asymptotic) matrix R_B of the model for B, and the fact that E is simply equal to minus the (generalized) inverse of the top-left submatrix in the (generalized) inverse of R_B. However, for the second part of Lemma 6, it is preferable to note that, from (17) and Lemma 3, we have:

$$E = -\frac{\partial\beta' Z'_B}{\partial\alpha}\left[\Omega_B - \Omega_B Z_B (Z'_B\,\Omega_B\,Z_B)^{-1} Z'_B\Omega_B\right]\frac{\partial Z_B\beta}{\partial\alpha'}.$$

The matrix Ω_B, as defined by (A.7), is positive semi-definite, i.e., for any vector V of R^J:

$$V'\Omega_B V \geq 0,$$

where the equality holds if and only if $V = \mu U_J$ for some scalar μ.

We shall show that the matrix Q in square brackets is also

positive semi-definite. Let us note that Ω_B defines a scalar product on any linear manifold that does not contain U_J. Hence, any vector V of R_J can be decomposed into the sum of a vector proportional to U_J, a vector in span Z_B and a vector in the orthogonal complement of span Z_B with respect to the scalar product Ω_B, i.e.:

$$V = \mu U_J + Z_B\lambda + W$$

for some scalar μ, some vector λ in R^a, and some vector W in $(\text{span } Z_B)^{\perp}$. It is now easy to show that

$$V'QV = W'\Omega_B W$$

Since Ω_B is positive semi-definite, then Q is positive semi-definite, and E is negative semi-definite.

To prove the second part of Lemma 6, Let us partition the matrix Λ into μ' and Λ_o where μ' is the first row of Λ. Thus

$$M_B\Lambda = U_J\mu' + Z_B\Lambda_o$$

Hence, if (A.32) holds, then

$$E = -\Lambda'_o\,Z'_B\,Q\,Z_B\,\Lambda_o$$

which is equal to zero.

Conversely, if (A.32) does not hold, then at least one column vector of $\frac{\partial Z_B \beta}{\partial \alpha'}$ does not belong to the space M_B, so that this column-vector has a non-null vector W in the decomposition introduced above. Therefore the norm of this column-vector with respect to Q, (which is also the norm of W with respect to Ω_B), is not null. Hence E cannot be the null matrix.

Q. E. D.

The condition (A.32) can be interpreted as follows: all the column-vectors of the matrix $\frac{\partial Z_B \beta}{\partial \alpha'}$ belong to the model space M_B spanned by U_J and the column-vectors of Z_B.

2.3. Proof of Corollary 1

To show that $\sum^1 - \sum^M$ is positive semi-definite, it suffices to show that $\left[\sum^M\right]^{-1} - \left[\sum^1\right]^{-1}$ is positive semi-definite. But from (14), (16), and (17), we have:

$$\left[\sum^M\right]^{-1} - \left[\sum^1\right]^{-1} = - \begin{bmatrix} E^o & 0 \\ 0 & 0 \end{bmatrix}$$

where E^o is the value of E at $(\alpha^o, \beta^o, Pr^o(B))$. Since E is negative semi-definite (Lemma 6), the first part of Corollary 1 is proved.

Also, it is clear that $\sum^1 = \sum^M$ if and only if $E^o = 0$. From the second part of Lemma 5, it follows that the two-step estimator is efficient if and only if (18) holds at (α^o, β^o).

2.4. Proof of Corollary 2

The efficiency part of the statement can be proved by showing that E = 0 for all $(\alpha, \beta, Pr(B))$, and by using Corollary 1. The equality between the two-step estimator and the FIML estimator requires, however, a direct proof.

For any given α, finding the β that maximizes $L_B(\alpha, \beta, f_B)$ is actually equivalent to finding the distribution Pr(B) that maximizes

$$L_B = f_B' \cdot \log Pr(B)$$

subject to the constraint

$$\log Pr(B) \in M_B(\alpha).$$

Since $M_B(\alpha)$, and then the set over which L_B is maximized do not depend on α, then the maximum value attained by L_B in that set does not depend on α. Hence:

$$L_B(\hat{\alpha}^M, \hat{\beta}^M, f_B) = L_B(\hat{\alpha}^1, \hat{\beta}^1, f_B).$$

Since $(\hat{\alpha}^M, \hat{\beta}^M)$ are FIML estimators, then because of the decomposition (13) we must have:

$$L_A(\hat{\alpha}^M, f_{AB}) + L_B(\hat{\alpha}^M, \hat{\beta}^M, f_B) \geq L_A(\hat{\alpha}^1, f_{AB}) + L_B(\hat{\alpha}^1, \hat{\beta}^1, f_B)$$

which, with the previous equation, implies:

$$L_B(\hat{\alpha}^M, f_{AB}) \geq L_A(\hat{\alpha}^1, f_{AB})$$

Since $\hat{\alpha}^1$ is the estimator obtained by maximizing L_A, then by the uniqueness of the optimizing solution, we must have:

$$\hat{\alpha}^1 = \hat{\alpha}^M$$

so that

$$\hat{\beta}^1 = \hat{\beta}^M.$$

3. Proofs of Results of Section 3

3.1 Proof of Theorem 2

It suffices to show that if $\hat{\delta}^r$ is an estimator of δ^o such that

$$\hat{P}r^r(B) = Pr^o(B) + O(T^{-1/2})$$

(which is satisfied by any initial estimator $\hat{\delta}^o$ that satisfied (21)), then the estimator $\hat{\delta}^{r+1}$ obtained at the next iteration of the natural iterative sequential procedure satisfies:

(i) $\hat{P}r^{r+1}(B) = Pr^o(B) + O(T^{-1/2})$

(ii) $\sqrt{T}(\hat{\delta}^{r+1} - \delta^o) \overset{D}{\to} N(0, \sum^1)$

where $\sum^1$ is the asymptotic covariance matrix of the two-step estimator studied in Section 2.

The normal equations for $\hat{\delta}^{r+1}$ are :

$$\left.\frac{\partial L_A}{\partial \alpha}\right|_{(\hat{\alpha}^{r+1},\ \hat{f}^r_{AB})} = 0$$

$$\left.\frac{\partial L_B}{\partial \beta}\right|_{(\hat{\alpha}^{r+1},\ \hat{\beta}^{r+1},\ f_B)} = 0 \qquad \text{(A.33)}$$

where $\hat{f}^r_{AB}$ is given by (20). Since $\hat{P}r(B) - Pr^o(B)$ is $O(T^{-1/2})$, it follows from (20) that $\hat{f}^r_{AB} - Pr^o(A, B)$ is also $O(T^{-1/2})$. Thus $\hat{f}^r_{AB}$ is a consistent estimator of $Pr^o(A, B)$.

Using the Implicit Function Theorem as in the proof of Theorem 1, it can readily be shown that $\hat{\alpha}^{r+1} - \alpha^o$ and $\hat{\beta}^{r+1} - \beta^o$ are both $O(T^{-1/2})$. Hence

$$\hat{P}r^{r+1}(B) = Pr^o(B) + O(T^{-1/2}).$$

Moreover $\hat{\alpha}^{r+1}$ and $\hat{\beta}^{r+1}$ are consistent estimators of α^o and β^o.

Taking a Taylor expansion around $(\alpha^o, Pr^o(A, B))$ of the first equation of (A.33) and noting that this value also satisfies that equation, we obtain:

$$0 = \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}(\hat{\alpha}^{r+1} - \alpha^o) + \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}(\hat{f}^r_{AB} - Pr^o(A, B)) + O(T^{-1}) \qquad \text{(A.34)}$$

brackets is null for the reason mentioned above. Since $\Omega_{A|B}$ and $D(Pr(B) \otimes U_I)$ are both block-diagonal, we can reverse the order of the matrix multiplication. This establishes (A.20).

Q. E. D.

From the previous lemmas, we have in particular:

<u>Lemma 5:</u>

At $(\alpha, \beta, Pr(A, B))$:

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \Omega \frac{\partial^2 L_A}{\partial f_{AB}\partial\alpha'} = - \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} \tag{A.21}$$

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \Omega \frac{\partial^2 L_B}{\partial f_{AB}\partial\alpha'} = 0 \tag{A.22}$$

$$\frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} \Omega \frac{\partial^2 L_B}{\partial f_{AB}\partial\beta'} = 0 \tag{A.23}$$

$$\frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} \Omega \frac{\partial^2 L_B}{\partial f_{AB}\partial\alpha'} = - \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} \tag{A.24}$$

$$\frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} \Omega \frac{\partial^2 L_B}{\partial f_{AB}\partial\beta'} = - \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \tag{A.25}$$

$$\frac{\partial^2 I_B}{\partial\beta\partial f'_{AB}} \Omega \frac{\partial^2 L_B}{\partial f_{AB}\partial\beta'} = - \frac{\partial^2 L_B}{\partial\beta\partial\beta'} \tag{A.26}$$

proof:

Straightforward from Lemmas 3 and 4.

Q. E. D.

Other lemmas which are more specific to each result of Section 2 - 5 are given below.

<u>2. Proofs of Results of Section 2</u>

2.1. Proof of Theorem 1

The likelihood function $L_A(\alpha, \beta, f_{AB})$ is strictly concave with respect to α, while the likelihood function $L_B(\alpha, \beta, f_B)$ is strictly concave with respect to β for any given α (see, e.g., S. J. Haberman (1974)). The normal equations for the two-step estimator $\hat{\delta}^1 = (\hat{\alpha}^{1\prime}, \hat{\beta}^{1\prime})'$ are:

$$\begin{aligned} \frac{\partial L_A}{\partial\alpha}\Big|_{\hat{\alpha}^1,\ f_{AB}} &= 0 \\ \frac{\partial L_B}{\partial\beta}\Big|_{(\hat{\alpha}^1,\ \hat{\beta}^1,\ f_B)} &= 0 \end{aligned} \tag{A.27}$$

These equations are also satisfied by $(\alpha^o, Pr^o(A, B))$ and $(\alpha^o, \beta^o, Pr^o(B))$. Since L_A has continuous second partial derivatives, then by the implicit function theorem, $\hat{\alpha}^1$ is a continuous function of

Now, since $\hat{Pr}^r(B) - Pr^o(B)$ and $f_B - Pr^o(B)$ are both $O(T^{-1/2})$, then :

$$\hat{Pr}^r(B) = f_B + O(T^{-1/2}).$$

Hence, from (3.15), we have:

$$\hat{f}^r_{AB} = D(f_{A|B}) \left[(f_B + O(T^{-1/2})) \otimes U_I\right]$$
$$= f_{AB} + V_{AB}$$

where

$$V_{AB} = D(f_{A|B})\ N_B\ O(T^{-1/2})$$

since $f_{A|B}$ is a consistent estimator of $Pr^o(A|B)$, i.e., $f_{A|B} = Pr^o(A|B) + o(1)$, it follows that

$$f_{AB} = D(Pr^o(A|B))\ N_B\ O(T^{-1/2}) + o(T^{-1/2})$$

Moreover, using (A.10), (A.12), we have:

$$\frac{\partial^2 L_A}{\partial \alpha \partial f'_{AB}} D(Pr^o(A|B))N_B = Z'_A\ \Omega^o_{A|B}\ N_B$$
$$= 0$$

where the second equation results from the fact that the column vectors of N_B belong to the kernel of $\Omega^o_{A|B}$. Equation (A.30) now becomes:

$$0 = \frac{\partial^2 L_A}{\partial \alpha \partial \alpha'} (\hat{\alpha}^{r+1} - \alpha^o) + \frac{\partial^2 L_A}{\partial \alpha \partial f'_{AB}} (f_{AB} - Pr^o(A, B))$$

$$+ O(T^{-1/2}) + O(T^{-1}) \tag{A.35}$$

Comparing (A.35) to (A.28), it follows that

$$\sqrt{T}(\hat{\alpha}^{r+1} - \alpha^o) = \sqrt{T}(\hat{\alpha}^1 - \alpha^o) + o(1).$$

Thus $\hat{\alpha}^{r+1}$ has the same asymptotic distribution as the two-stage estimator $\hat{\alpha}^1$. To show that $\hat{\beta}^{r+1}$ and $\hat{\beta}^1$ have the same asymptotic distribution, it suffices to compare the Taylor expansion of the second equation of (A.33) to Equation (A.29), and to use the fact that $\hat{\alpha}^{r+1}$ and $\hat{\alpha}^1$ have the same asymptotic distribution.

3.2. Two Lemmas

To show Corollary 3, we shall need the first partial derivatives of the mapping

$$f^N(\alpha, \beta, f_{A|B}) = D(f_{AB})\ [Pr(B) \otimes U_I] \tag{A.36}$$

associated with the adjustment (20).

Lemma 7:

The first partial derivatives of $f^N(\alpha, \beta, f_{AB})$ evaluated at $(\alpha, \beta, Pr(A, B))$ are:

$$\frac{\partial f^N}{\partial \alpha'} = D(Pr(A|B))\left[(\Omega_B \frac{\partial Z_B \beta}{\partial \alpha'}) \otimes U_I\right] \tag{A.37}$$

$$\frac{\partial f^N}{\partial \beta'} = D(Pr(A|B))\left[(\Omega_B Z_B) \otimes U_I\right] \tag{A.38}$$

$$\frac{\partial f^N}{\partial f'_{AB}} = I - D(Pr(A|B))(N'_B \otimes U_I) \tag{A.39}$$

proof:

Equation (A.36) is equivalent to:

$$f^N(\alpha, \beta, f_{AB}) = D(Pr(B) \otimes U_I)\ f_{A|B}$$

Equation (A.39) now follows from:

$$\frac{\partial f_{A|B}}{\partial f'_{AB}} = D(f_B^{-1} \otimes U_I)\ [I - D(f_{A|B})(N'_B \otimes U_I)].$$

Equations (A.37) and (A.38) directly follows by differentiating Equation (A.36) and by using the formulae for the partial derivatives of Pr(B) with respect to α and β which are given in the proof of Lemma 3.

Q. E. D.

A useful property of the right-hand side of (A.39) is:

Lemma 8:

The (non-symmetric) matrix (A.39) is idempotent, i.e.,:

$$[I - D(Pr(A|B))\ (N'_B \otimes U_I)]^2 = I - D(Pr(A|B))\ (N'_B \otimes U_I) \tag{A.40}$$

proof:

It suffices to show that

$$[D(Pr(A|B))\ (N'_B \otimes U_I)]^2 = D(Pr(A|B))\ (N'_B \otimes U_I)$$

or equivalently

$$(N'_B \otimes U_I)\ D(Pr(A|B))\ (N'_B \otimes U_I) = N'_B \otimes U_I.$$

we have:

$$N'_B \otimes U_I = N_B N'_B.$$

and

$$N'_B\ D(Pr(A|B))\ N_B = I$$

The desired equality follows.

Q. E. D.

3.3. Proof of Corollary 3

In order to determine the order in probability of the remainder in a Taylor expansion we shall first show that $\hat{\delta}^L - \delta^o$ is $O(T^{-1/2})$. Since $f_{AB} - Pr^o(A, B))$ is $O(T^{-1/2})$, this will be established by showing that $\hat{\delta}^L$ is a continuous function of f_{AB} that has a first derivative at $Pr^o(A, B)$.

The estimator $\hat{\delta}^L$ satisfies:

$$\left.\frac{\partial L_A}{\partial \alpha}\right|_{(\hat{\alpha}^L,\ f^N(\hat{\alpha}^L,\ \hat{\beta}^L, f_{AB}))} = 0$$
$$\left.\frac{\partial L_B}{\partial \beta}\right|_{(\hat{\alpha}^L,\ \hat{\beta}^L,\ f_B)} = 0 \tag{A.41}$$

since $\hat{\delta}^L$ is consistent, then $\hat{f}^L_{AB}$ is a consistent estimator of $Pr^o(A, B)$ (see (23)). Thus $(\hat{\alpha}^L, \hat{f}^L_{AB})$ and $(\hat{\alpha}^L, \hat{\beta}^L, f_B)$ converge in probability to $(\alpha^o, Pr^o(A, B)$ and $(\alpha^o, \beta^o, Pr^o(B))$. Moreover $\frac{\partial L_A}{\partial \alpha}$, $\frac{\partial L_B}{\partial \beta}$, and f^N are continuously differentiable. Hence the fact that $\hat{\delta}^L$ is a function of f_{AB} with the above required properties simply results from the Implicit Function Theorem provided the matrix

$$Q = \begin{bmatrix} \dfrac{\partial^2 L_A}{\partial \alpha \partial \hat{\alpha}^{L\prime}} & \dfrac{\partial^2 L_A}{\partial \alpha \partial \hat{\beta}^{L\prime}} \\[2ex] \dfrac{\partial^2 I_B}{\partial \beta \partial \hat{\alpha}^{L\prime}} & \dfrac{\partial^2 I_B}{\partial \beta \partial \hat{\beta}^{L\prime}} \end{bmatrix}$$

is non singular at $(\alpha^o, \beta^o, Pr^o(A, B))$.

We have :

$$\frac{\partial^2 L_A}{\partial \alpha \partial \hat{\alpha}^{L\prime}} = \frac{\partial^2 L_A}{\partial \alpha \partial \alpha'} + \frac{\partial^2 L_A}{\partial \alpha \partial f'_{AB}} \frac{\partial f^N}{\partial \alpha'} \tag{A.42}$$

$$\frac{\partial^2 L_A}{\partial \alpha \partial \hat{\beta}^{L\prime}} = \frac{\partial^2 L_A}{\partial \alpha \partial f'_{AB}} \frac{\partial f^N}{\partial \beta'} \tag{A.43}$$

$$\frac{\partial^2 L_B}{\partial \beta \partial \hat{\alpha}^{L\prime}} = \frac{\partial^2 L_B}{\partial \beta \partial \alpha'} \tag{A.44}$$

$$\frac{\partial^2 I_B}{\partial \beta \partial \hat{\beta}^{L\prime}} = \frac{\partial^2 L_B}{\partial \beta \partial \beta'} \tag{A.45}$$

where all the derivatives are evaluated at $(\alpha^o, \beta^o, Pr^o(A, B))$. But from (A.10) and Lemmas 3 and 7, we have:

$$\frac{\partial^2 L_A}{\partial \alpha \partial f'_{AB}} \frac{\partial f^N}{\partial \alpha'} = Z'_A\ \Omega^o_{A|B} \left[(\Omega_B \frac{\partial Z_B \beta}{\partial \alpha'}) \otimes U_I \right] \tag{A.46}$$

$$= 0$$

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}\frac{\partial f^N}{\partial\beta'} = Z'_A\ \Omega^o_{A|B}\left[(\Omega_B Z_B)\otimes U_I\right].ne7 \tag{A.47}$$

$$= 0$$

where the second equality of either (A.46) or (A.47) follows from the fact that the vectors of the form $V\otimes U_I$ belong to the kernel of $\Omega_{A|B}$. Thus

$$Q = \begin{bmatrix} \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} & 0 \\ \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} & \frac{\partial^2 L_B}{\partial\beta\partial\beta'} \end{bmatrix}$$

which is non-singular.

Taking now the Taylor expansion of (A.41) around $(\alpha^o, \beta^o, Pr^o(A, B))$ and using Equations (A.42) - (A.44) giving the partial derivatives, and Equations (A.46) - (A.47), we obtain:

$$0 = \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}(\hat{\alpha}^L - \alpha^o) + \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}\frac{\partial f^N}{\partial f'_{AB}}(f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.48}$$

$$0 = \frac{\partial^2 L_B}{\partial\beta\partial\alpha'}(\hat{\alpha}^L - \alpha^o) + \frac{\partial^2 L_B}{\partial\beta\partial\beta'}(\hat{\beta}^L - \beta^o) + \frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}}(f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.49}$$

Moreover, from (A.12), (A.37), and (A.49), it follows that

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}\frac{\partial f^N}{\partial f'_{AB}} = \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}.$$

It now suffices to compare the resulting equations (A.48) - (A.49) to the equations (A.28) - (A.29) in order to establish that $\hat{\delta}^L$ and $\hat{\delta}^1$ have the same asymptotic distribution.

4. Proofs of Results of Section 4

4.1. Some Lemmas

We shall need the partial derivatives of the mapping

$$f^E(\alpha, \beta, f_{AB}) = f_{AB} + D(f_{AB})Z_A\left[\frac{\partial^2\hat{L}_A}{\partial\alpha\partial\alpha'}\right]^{-1}\frac{\partial L_B}{\partial\alpha}\Big|_{(\alpha,\ \beta,\ f_B)} \tag{A.50}$$

where the matrix in bracket is some consistent estimate of the second partial derivatives of L_A at $(\alpha^o, Pr^o(A,B))$.

Lemma 9

The first partial derivatives of $f^E(\alpha, \beta, f_{AB})$ evaluated at $(\alpha, \beta, Pr(A, B))$ are:

$$\frac{\partial f^E}{\partial\alpha'} = -D(Pr(A, B))\ Z_A\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}\frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} \tag{A.51}$$

$$\frac{\partial f^E}{\partial\beta'} = -D(Pr(A, B))\ Z_A\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}\frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \tag{A.52}$$

$$\frac{\partial f^E}{\partial f'_{AB}} = I - D(Pr(A, B))\ Z_A \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} \frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} \qquad (A.53)$$

proof:

Since $\frac{\partial^2 \hat{L}_A}{\partial\alpha\partial\alpha'}$ is an estimator, it is a function A (say) of f_{AB}, i.e.,

$$\frac{\partial^2 \hat{L}_A}{\partial\alpha\partial\alpha'} = A(f_{AB}).$$

Since $\frac{\partial^2 \hat{L}_A}{\partial\alpha\partial\alpha'}$ is a consistent estimator of $\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}$ evaluated at $(\alpha, Pr(A, B))$, and since f_{AB} is a consistent estimator of Pr(A, B), it follows by the unicity of the probability limit that:

$$A(Pr^o(A, B)) = \left.\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right|_{(\alpha,\ Pr(A,\ B))}$$

Let X be the matrix by which $\frac{\partial L_B}{\partial\alpha}$ is premultiplied in (A.50). The matrix X depend only on f_{AB}. Thus $\frac{\partial L_B}{\partial\alpha}$ is the only term in (A.50) that depends on α and β. Equations (A.51) - (A.52) straightforwardly follows from the above remark.

To prove Equation (A.53), it suffices to note that:

$$\frac{\partial f^E}{\partial f'_{AB}} = I + X \frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} + \frac{\partial X}{\partial f'_{AB}} (U_{IJ} \otimes \frac{\partial L_B}{\partial\alpha})$$

where $\frac{\partial X}{\partial f'_{AB}}$ is the IJ $\times$ aIJ matrix of which the (i,j)-th block is the IJ $\times$ a matrix $\frac{\partial X}{\partial f_{ij}}$. From (25) it follows that, at $(\alpha, \beta, Pr(A, B))$, $\frac{\partial L_B}{\partial\alpha} = 0$. Hence (A.53) follows.

Q. E. D.

The premultiplication of each of the above first derivatives by $\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}}$ is also useful.

<u>Lemma 10</u>

At $(\alpha, \beta, Pr(A, B))$, we have:

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \frac{\partial f^E}{\partial\alpha'} = \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} \qquad (A.54)$$

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \frac{\partial f^E}{\partial\beta'} = \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \qquad (A.55)$$

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \frac{\partial f^E}{\partial f'_{AB}} = \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} + \frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} \qquad (A.56)$$

proof:

From (A.12), we have:

$$\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} D(Pr(A, B))\ Z_A = Z'_A\ [I - D(Pr(A|B))N_B N'_B]\ D(Pr(A, B))\ Z_A$$

$$= Z_A' \, \Omega_{A|B} \, D(Pr(B) \otimes U_I) \, Z_A$$

$$= - \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}$$

Where the first equality follows from $N_B \otimes U_I = N_B N_B'$, the second equality from $D(Pr(A, B)) = D(Pr(A|B)) \, D(Pr(B) \otimes U_I)$ and (A.10), and the third equality from (A.8), (A.11) and the fact that $\Omega_{A|B}$ and $D(Pr(B) \otimes U_I)$ are both block diagonal so that we can reverse the order of the matrix multiplication. Equations (A.54) – (A.56) straightforwardly follow from Lemma 8.

Q. E. D.

The next Lemma considers a sequence of n-dimensional (non-random) vectors y_r that satisfy:

$$0 = u + P\, y_{r-1} + Q\, y_r \quad \text{for} \quad r \geq 2 \tag{A.57}$$

for some n-dimensional vector u, and some $n \times n$ matrices P and Q.

<u>Lemma 11:</u>

Suppose that Q and P + Q are non-singular. Then for any given initial vector y_1, the solution y_r, $r \geq 1$, of Equation (A.57) is:

$$y_r = -(P + Q)^{-1}\left[u + (-1)^r X^{r-1}\{u + (P + Q) y_1\}\right] \tag{A.58}$$

where $X = PQ^{-1}$.

proof:

Let R.H.S. be the right-hand side of (A.57). Using (A.58), it follows that

$$(-1)^t \text{ RHS} = \left[P(P + Q)^{-1} X^{r-2} - Q(P + Q)^{-1} X^{r-1}\right]\{u + (P + Q) y_1\}$$

But:

$$\begin{aligned} P(P + Q)^{-1} - Q(P + Q)^{-1} X &= I - Q(P + Q)^{-1}(I + X) \\ &= I - Q(P + Q)^{-1}(P + Q) Q^{-1} \\ &= 0 \end{aligned}$$

This establishes the desired result.

Q. E. D.

We shall let P, Q, u, v (where $v = -Q^{-1} y_1$) be the following matrices and vectors:

$$P = \begin{bmatrix} \dfrac{\partial^2 L_B}{\partial\alpha\partial\alpha'} & ; & \dfrac{\partial^2 L_B}{\partial\alpha\partial\beta'} \\ 0 & ; & 0 \end{bmatrix} \quad ; \quad Q = \begin{bmatrix} \dfrac{\partial^2 L_A}{\partial\alpha\partial\alpha'} & ; & 0 \\ \dfrac{\partial^2 L_B}{\partial\beta\partial\alpha'} & ; & \dfrac{\partial^2 L_B}{\partial\beta\partial\beta'} \end{bmatrix} \tag{A.59}$$

$$u = \begin{bmatrix} \dfrac{\partial^2 L_A}{\partial\alpha\partial f_{AB}'} + \dfrac{\partial^2 L_B}{\partial\alpha\partial f_{AB}'} \\ \dfrac{\partial^2 L_B}{\partial\beta\partial f_{AB}'} \end{bmatrix} (f_{AB} - Pr^o(A, B)) \quad ;$$

$$v = \begin{bmatrix} \dfrac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \\ \\ \dfrac{\partial^2 L_B}{\partial\beta\partial f'_{AB}} \end{bmatrix} (f_{AB} - Pr^o(A,\ B)).$$

(A.60)

where all the partial derivatives are evaluated at $(\alpha^o,\ \beta^o,\ Pr^o(A,\ B))$.

Let us note that P + Q is simply equal to the hessian of the complete likelihood function $L(\alpha,\ \beta,\ f_{AB})$ evaluated at $(\alpha^o,\ \beta^o,\ Pr^o(A,\ B))$ (see (13)). The previous lemma requires that P+Q be non singular. This condition is satisfied since $\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}$ and $F = \dfrac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + E$ are both non singular (as we shall see in the next Lemma). Thus one can apply the formula for the inverse of a partitioned matrix:

$$(P + Q)^{-1} =$$

$$\begin{bmatrix} F^{-1} & ; & -F^{-1}\dfrac{\partial^2 L_B}{\partial\alpha\partial\beta'}\left[\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \\ \\ -\left[\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}\dfrac{\partial^2 L_B}{\partial\beta\partial\alpha'}F^{-1} & ; & \left[\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} + \left[\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}\dfrac{\partial^2 L_B}{\partial\beta\partial\alpha'}F^{-1}\dfrac{\partial^2 L_B}{\partial\alpha\partial\beta'}\left[\dfrac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \end{bmatrix}$$

(A.61)

Theorem 3 involves the a X a matrices F and G which are defined by Equations (26) and (27). The next Lemma gives some properties of these matrices. Let K_r be the a X a matrix:

$$K_r = G^r \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} G'^r + G^{r-1} E\, G'^{r-1} \qquad \text{(A.62)}$$

where E is defined in (17), and all the partial derivatives are evaluated at $(\alpha,\ \beta,\ Pr(A,\ B))$.

<u>Lemma 12:</u>

For any $r \geq 1$, the matrix $I + G^{2r-1}$ is nonsingular, and the matrices F, $F + K_r$, $(I + G^{2r-1})F$ are symmetric negative definite. Moreover, we have:

$$FG'^{r-1} = G^{r-1}F \qquad \text{(A.63)}$$

and

$$F(F + K_r)^{-1}F = (I + G^{2r-1})^{-1}F \qquad \text{(A.64)}$$

proof:

We have:

$$I + G^{2r-1} = \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + G^{r-1}E\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} G^{r-2}E\right]\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}$$

$$= \left[\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right] + G^{r-1}E\, G'^{r-1}\right]\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}$$

The matrix in brackets is negative definite since it is the sum of a negative definite matrix and a negative semi-definite matrix (see Lemma 6). Hence $I + G^{2r-1}$ is nonsingular.

Also, from (26) and Lemma 6, it follows that F is negative definite. Since K_r is clearly negative semi-definite, then $F + K_r$ is negative definite.

To establish (A.63), it suffices to note that:

$$FG'^{r-1} = \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\left[\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} E\right]^{r-1} + E\left[\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} E\right]^{r-1}$$

$$= \left[E\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}\right]^{r-1} \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + \left[E\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1}\right]^{r-1} E$$

$$= G^{r-1}F$$

where we have used the symmetry of E.

To establish (A.64) it suffices to show that:

$$F + K_r = (I + G^{2r-1})F$$

which straightforwardly follows from (A.62) and

$$K_r = G^{r-1}\left[G\,\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\,G + E\right] G'^{r-1}$$

$$= G^{r-1}E\left[\left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} E + I\right]G'^{r-1}.\text{ne4}$$

$$= G^{r}FG'^{r-1}$$

Finally, Equation (A.64) shows that $(I + G^{2r-1})F$ is negative definite.

Q. E. D.

Finally, we shall need the asymptotic distribution of the vector $[u', w']'$, where u is defined by (A.60), and w by:

$$w = \left[-G\,\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} + \frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}} - \frac{\partial^2 L_B}{\partial\alpha\partial\beta'}\left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}\frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}}\right](f_{AB} - Pr^{o}(A, B)) \tag{A.65}$$

where all the partial derivatives are again evaluated at $(\alpha^{o}, \beta^{o}, (A, B))$.

<u>Lemma 13:</u>

$$\sqrt{T}\begin{bmatrix} u \\ w \end{bmatrix} \xrightarrow{D} N(0, C)$$

where

$$C = \begin{bmatrix} C_{uu} & 0 \\ 0 & C_{ww} \end{bmatrix} \tag{A.66}$$

$$C_{uu} = -(P + Q) \quad ; \quad C_{ww} = -\left[G \frac{\partial^2 L_a}{\partial a \partial'} G' + E\right]$$

proof:

The convergence to a normal distribution directly follows from (A.60), (A.65), and Lemma 1. To derive the asymptotic covariance matrix, it suffices to apply Lemmas 1 and 5, and the definitions of E and G (Equations (17) and (27)).

Q. E. D.

4.2. Proof of Theorem 3

Given the estimator $\hat{\delta}^{r-1}$ obtained at the r-1 iteration, the estimator $\hat{\delta}^{r}$ obtained at the next iteration satisfied the normal equations.

$$\left.\frac{\partial L_A}{\partial \alpha}\right|_{(\hat{\alpha}^r,\ \hat{f}_{AB}^{r-1})} = 0$$
$$\left.\frac{\partial L_B}{\partial \beta}\right|_{(\hat{\alpha}^r,\ \hat{\beta}^r,\ f_B)} = 0 \tag{A.67}$$

where $\hat{f}_B^{r-1}$ is now equal to $f^E(\hat{\alpha}^{r-1},\ \hat{\beta}^{r-1},\ f_{AB})$.

Let us suppose that the estimator $\hat{\delta}^{r-1}$ is consistent and that $\hat{\delta}^{r-1} - \delta^o$ is $O(T^{-1/2})$. This is certainly satisfied by the estimator $\hat{\delta}^1$ obtained at the first iteration since $\hat{\delta}^1$ is simply the two-step estimator of Section 2. Moreover, since

$$f^E(\alpha^o,\ \beta^o,\ Pr^o(A,\ B)) = Pr^o(A,\ B) \tag{A.68}$$

(see Equation (25)), it follows that $\hat{f}_{AB}^{r-1}$ is a consistent estimator of $Pr^o(A,\ B)$, and that $\hat{f}_{AB}^{r-1} - Pr^o(A,\ B)$ is $O(T^{-1/2})$. Now, $(\alpha^o,\ Pr^o(A,\ B))$ and $(\alpha^o,\ \beta^o,\ Pr^o(B))$ satisfy Equations (A.54). From the Implicit Function Theorem, it follows that $\hat{\delta}^r = (\hat{\alpha}^{r\prime},\ \hat{\beta}^{r\prime})'$ is a consistent estimator of δ^o, and that $\hat{\delta}^r - \delta^o$ is $O(T^{-1/2})$.

Taking the Taylor expansions of Equations (A.67) around

(α^o, $Pr^o(A, B)$) and (α^o, β^o, $Pr^o(B)$), and using the fact that these values satisfy (A.54), we obtain, for $r \geq 2$:

$$0 = \frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} \left[\frac{\partial f^E}{\partial\alpha'} (\hat{\alpha}^{r-1} - \alpha^o) + \frac{\partial f^E}{\partial\beta'} (\hat{\beta}^{r-1} - \beta^o) + \frac{\partial f^E}{\partial f'_{AB}} (f_{AB} - Pr^o(A, B))\right]$$

$$+ \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} (\hat{\alpha}^r - \alpha^o) + O(T^{-1}) \tag{A.69}$$

$$0 = \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} (\hat{\alpha}^r - \alpha^o) + \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} (\hat{\beta}^r - \beta^o)$$

$$+ \frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}} (f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.70}$$

where all the partial derivatives are evaluated at (α^o, β^o, $Pr^o(A, B)$). On the other hand, for $r = 1$, the Taylor expansions are given by (A.28) - (A.29).

From Lemma 10, it follows that the system (A.63) - (A.70) for $r \geq 2$ can be written as the vector first difference equation:

$$0 = u + P(\hat{\delta}^{r-1} - \delta^o) + Q(\hat{\delta}^r - \delta^o) + O(T^{-1}) \tag{A.71}$$

where u, P, and Q are defined by (A.59) - (A.60). For $r = 1$, the Taylor expansions (A.28) - (A.29) are:

$$0 = v + Q(\hat{\delta}^1 - \delta^o) + O(T^{-1}). \tag{A.72}$$

where v is defined in (A.60).

From Lemma 11, it follows that for $r \geq 1$:

$$\hat{\delta}^r - \delta^o = -(P + Q)^{-1}\left[u + (-1)^r X^{r-1}(u - v - Xv)\right] + O(T^{-1}) \tag{A.73}$$

where $X = PQ^{-1}$. Since Q is triangular, its inverse is readily derived.

$$Q^{-1} = \begin{bmatrix} \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} & ; & 0 \\ -\left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]^{-1} & ; & \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \end{bmatrix} \tag{A.74}$$

Hence X can readily be computed. Since the bottom submatrices of P, and hence of X, are null, the matrix X^r can straightforwardly be derived. We have for $r \geq 1$:

$$X^r = \begin{bmatrix} G^r & ; & G^{r-1} \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \\ 0 & ; & 0 \end{bmatrix} \tag{A.75}$$

From (A.60) and (A.75), it is readily seen that:

$$u - v - Xv = \begin{bmatrix} w \\ 0 \end{bmatrix} \qquad (A.76)$$

where w is as defined by (A.65). Hence (A.73) becomes:

$$\hat{\delta}^r - \delta^o = -(P + Q)^{-1}u - (P + Q)^{-1} L_{r-1}w \qquad (A.77)$$

where

$$L_{r-1} = \begin{bmatrix} G^{r-1} \\ 0 \end{bmatrix}.$$

From Lemma 13, it follows that $\sqrt{T}(\hat{\delta}^r - \delta^o)$ converges in distribution to a normal distribution with mean zero and some covariance matrix $\sum^r$ equal to:

$$\sum{}^r = -(P + Q)^{-1} - (P + Q)^{-1}L_{r-1}\left[G \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} G' + E\right] L'_{r-1}(P + Q)^{-1} \qquad (A.78)$$

Using the formula (A.61) for the inverse of (P + Q), and after some matrix operations, we obtain:

$$\sum{}^r = \begin{bmatrix} \sum^r_{\alpha\alpha} & \sum^r_{\alpha\beta} \\ \sum^r_{\beta\alpha} & \sum^r_{\beta\beta} \end{bmatrix} \qquad (A.79)$$

where

$$\sum{}^r_{\alpha\alpha} = -F^{-1}\left[F + K_r\right]F^{-1}$$

$$\sum{}^r_{\beta\alpha} = \sum{}^r_{\alpha\beta}{}' = \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} F^{-1}\left[F + K_r\right]F^{-1}$$

$$\sum{}^r_{\beta\beta} = -\left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} - \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} F^{-1}[F + K_r]F^{-1} \frac{\partial^2 L_B}{\partial\alpha\partial\beta'}\left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}$$

To show that $\sum^r$ is equal to the right-hand side of (3.25), it now suffices to use the formula for the inverse of a partitioned matrix, and Equation (A.64).

4.3. Another Lemma.

Corollaries 4 and 5 involve the matrix G defined by (27). This matrix is not necessarily symmetric. Nevertheless, we have:

<u>Lemma 14:</u>

The $a \times a$ matrix G always has "a" real nonnegative roots which may or may not be distinct. The largest characteristic root λ_M satisfies:

$$\lambda_M = \max_{x \neq 0} \frac{x' E x}{x' \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} x} \qquad (A.80)$$

Moreover, the matrix G is similar to the diagonal matrix $D(\underline{\lambda})$ where $\underline{\lambda}$ is the vector of characteristic roots, i.e., there exists a non-singular matrix Z such that:

$$G = Z^{-1} D(\underline{\lambda}) Z \tag{A.81}$$

In addition the matrix Z satisfies:

$$Z E Z' = -D(\underline{\lambda}) \quad ; \quad Z \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} Z' = -I \tag{A.82}$$

i.e., diagonalizes simultaneously E and $\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}$.

proof:

The characteristic polynomial of G is:

$$\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}$$

which is clearly equivalent to:

$$\det\left[(-E) - \lambda \left[-\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]\right] = 0$$

This latter equation shows that the characteristic roots of G are the characteristic values of the regular pencil of positive quadratic forms $-E - \lambda \left[-\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]$ (see F. R. Gantmacher (1960, p. 310)). It follows that the a characteristic roots of G are real (Gantmacher (1960, Theorem 8, p. 310)). Since the smallest characteristic value

is equal to the minimum of the ratio $x'(-E)x/x'\left[-\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}\right]x$ (Gantmacher (1960, Theorem 10, p. 319)), and since this ratio is always nonnegative, it follows that all the characteristic roots of G are nonnegative. On the other hand, the largest characteristic of G must satisfy (A.80) (Gantmacher (1960, Theorem 13, p. 322)).

Finally, Equation (A.82) directly follows from Gantmacher (1960, Theorem 9, p. 314), after a sign change, while Equation (A.81) follows from (A.82) and (27) after straightforward matrix operations.

Q. E. D.

4.4. Proof of Corollary 4

From (3.11) and (A.59), we have:

$$\sum^M = -(P + Q)^{-1} \tag{A.83}$$

Since the matrix in brackets in (A.78) is clearly negative semi-definite, then $\sum^M - \sum^r$ is negative semi-definite.

To show that $\sum^{r+1} - \sum^r$ is negative semi-definite or negative definite, we shall show that $\left[\sum^r\right]^{-1} - \left[\sum^{r+1}\right]^{-1}$ is negative semi-definite or negative definite. From (28), we have

$$\left[\sum^r\right]^{-1} - \left[\sum^{r+1}\right]^{-1} = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}$$

where

$$S = -(I + G^{2r-1})^{-1}F + (I + G^{2r+1})^{-1}F$$
$$= F\Big[(F + K_{r+1})^{-1} - (F + K_r)^{-1}\Big]F$$

(The second equality follows from (A.64).) Hence S is negative semi-definite if and only if the matrix in brackets is negative semi-definite, or if and only if

$$(F + K_r) - (F + K_{r+1}) = K_r - K_{r+1}$$

is negative semi-definite. Now, from (A.62), we have:

$$K_r - K_{r+1} = G^{r-1}\left[E + G\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}G' - G\,E\,G' - G^2\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'}G'^2\right]G'^{r-1} \tag{A.84}$$

Moreover, from (A.81), we have:

$$G^r = Z^{-1}D(\underline{\lambda}^r)Z \tag{A.85}$$

where $\underline{\lambda}^r$ is the vector of which the components are the r-th powers of the characteristic roots.

From (A.82), (A.84), and (A.85) it follows that :

$$K_r - K_{r+1} = Z^{-1}[-D(\underline{\lambda}^{2r-1}) - D(\underline{\lambda}^{2r}) + D(\underline{\lambda}^{2r+1}) + D(\underline{\lambda}^{2r+2})]Z^{-1},$$
$$= Z^{-1}D(\underline{\lambda}^{2r+2} + \underline{\lambda}^{2r+1} - \underline{\lambda}^{2r} - \underline{\lambda}^{2r-1})Z^{-1},$$

Therefore $K_r - K_{r+1}$, and hence $\sum^{r+1} - \sum^{r}$ are negative semi-definite

if and only if all the elements of the diagonal matrix in the above equation are nonpositive. Since the typical element of this diagonal matrix is $\lambda_i^{2r-1}[\lambda_i^3 + \lambda_i^2 - \lambda_i - 1]$ which is equal to $\lambda_i^{2r-1}(\lambda_i + 1)(\lambda_i^2 - 1)$, and since all the characteristic roots λ_i are real and nonnegative (lemma 14), it follows that $\sum^{r+1} - \sum^{r}$ is negative semi-definite if and only if the largest characteristic root of G is not greater than one. It also follows that $\sum^{r+1} - \sum^{r}$ is negative definite if and only if the largest characteristic root of G is strictly less than one.

4.5. Proof of Corollary 5

We have

$$F + \frac{\partial^2 L_B}{\partial\alpha\partial\beta'}\left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1}\frac{\partial^2 L_B}{\partial\beta\partial\alpha'} = \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'}. \tag{A.86}$$

Then, the result is straightforward from (16), (28), and the fact that G can be diagonalized (lemma 14) so that (A.85) holds.

4.6. Proof of Corollary 6

Since $\hat{f}^L_{AB} = f^E(\hat{\alpha}^L, \hat{\beta}^L, f_{AB})$ (see Equations (A.50) and (32)), and since $\hat{\delta}^L$ is a consistent estimator of δ^o, then it follows from (A.68) that $\hat{f}^L_{AB}$ is a consistent estimator of $Pr^o(A, B)$. Moreover, applying the same argument as in the proof of Corollary 3, it follows that $\hat{\delta}^L - \delta^o$ is $O(T^{-1/2})$ provided the matrix

$$R = \begin{bmatrix} \frac{\partial^2 L_A}{\partial\alpha\partial\hat{\alpha}^{L\prime}} & ; & \frac{\partial^2 L_A}{\partial\alpha\partial\hat{\beta}^{L\prime}} \\ \frac{\partial^2 L_B}{\partial\beta\partial\hat{\alpha}^{L\prime}} & ; & \frac{\partial^2 L_B}{\partial\beta\partial\hat{\beta}^{L\prime}} \end{bmatrix}$$

is nonsingular at $(\alpha^o, \beta^o, Pr^o(A, B))$ where $\hat{\alpha}^L$ and $\hat{\beta}^L$ satisfy (A.41) with f^E instead of f^N.

The second partial derivatives in R satisfy Equations (A.42) - (A.45) with f^E instead of f^N. Then, using Lemma 10 we get:

$$R = \begin{bmatrix} \frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 L_B}{\partial\beta\partial\beta'} & ; & \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \\ \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} & ; & \frac{\partial^2 L_B}{\partial\beta\partial\beta'} \end{bmatrix}$$

which is equal to P + Q (see A.59). Hence R is nonsingular (see the discussion following (A.60)).

Taking now the Taylor expansion of (A.41) with f^E instead of f^N around $(\alpha^o, \beta^o, Pr^o(A, B))$ and using Lemma 10, we obtain:

$$0 = \left[\frac{\partial^2 L_A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 L_B}{\partial\alpha\partial\alpha'} + \right](\hat{\alpha}^L - \alpha^o) + \frac{\partial^2 L_B}{\partial\alpha\partial\beta'}(\hat{\beta}^L - \beta^o)$$

$$+ \left[\frac{\partial^2 L_A}{\partial\alpha\partial f'_{AB}} + \frac{\partial^2 L_B}{\partial\alpha\partial f'_{AB}}\right] (f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.87}$$

$$0 = \frac{\partial^2 L_B}{\partial\beta\partial\alpha'} (\hat{\alpha}^L - \alpha^o) + \frac{\partial^2 L_B}{\partial\beta\partial\beta'} (\hat{\beta}^L - \beta^o) + \frac{\partial^2 L_B}{\partial\beta\partial f'_{AB}} (f_{AB} - Pr^o(A, B)) + O(T^{-1}) \tag{A.88}$$

It is easy to see that these equations are similar to the Taylor expansion of the normal equations for the FIML estimator:

$$0 = \frac{\partial L_A}{\partial\alpha}(\hat{\alpha}^M, f_{AB}) + \frac{\partial L_B}{\partial\alpha} (\hat{\alpha}^M, \hat{\beta}^M, f_B)$$

$$0 = \frac{\partial L_B}{\partial\beta} (\hat{\alpha}^M, \hat{\beta}^M, f_B)$$

Hence $\hat{\delta}^L$ and $\hat{\delta}^M$ have the same asymptotic distribution, and $\hat{\delta}^L$ is asymptotically efficient.

5. Generalizations

We shall only consider the generalization to the case in which all individual characteristics are assumed to be qualitative. In order to see if the properties of the efficient iterative sequential procedure still hold, it suffices to extend the lemmas that are used in the proofs of these properties to this more general case.

Lemma 1 is modified as follows:

<u>Lemma 1':</u>

$$\text{plim } f_{AB|C} = Pr^{o}(A,B|C) \tag{A.89}$$

and

$$\sqrt{T}(f_{AB|C} - Pr^{o}(A,B|C)) \xrightarrow{D} N(0,\ V^{o}_{AB|C}) \tag{A.90}$$

where

$$V^{o}_{AB|C} = D(Pr^{o}(C) \otimes U_{IJ})^{-1}\ \Omega^{o}_{AB|C} \tag{A.91}$$

and $\Omega^{o}_{AB|C}$ is the block-diagonal matrix of which the k-th block is:

$$\Omega_{AB|k} = D(Pr(A,B|k)) - Pr(A,B|k)\ Pr(A,B|k)' \tag{A.92}$$

proof:

This directly follows from (i) the assumption plim $f_c = Pr^{o}(C)$, (ii) the fact that the samplings are independent for different values of C, (iii) the fact that Lemma 1 now holds for $f_{AB|k}$ for any k where T is now Tf_k, and (iv) the equation:

$$\sqrt{T}[f_{AB|k} - Pr^{o}(A,B|k)] = \frac{1}{\sqrt{f_k}} \sqrt{Tf_k}[f_{AB|k} - Pr^{o}(A,B|k)].ne5$$

Q. E. D.

Let us note that if $p^{o}_{k} = 0$ for some k, then the vectors in

(A.80) - (A.90) should accordingly be reduced by deleting the components associated with C = k.

All the Lemmas 2 - 5, and 7 - 10 continue to hold provided they are written for C = k. Specifically, in the statements of these lemmas, a subscript k is now attached to the symbols L_A, L_B, Z_A, Z_B, f^N, and f^E, (e.g., L_A becomes L_{Ak}), while the other symbols are defined by conditioning on C = k (e.g., f_{AB}, $Pr(A|B)$, $\Omega_{A|B}$, and Ω become respectively $f_{AB|k}$, $Pr(A|B,k)$, $\Omega_{A|B,k}$, and $\Omega_{AB|k}$).

The matrices E, F, G, P and Q are defined as before (see Equations (17), (26), (27), and (A.59)), and thus involve the log-likelihoods L_A and L_B (not L_{Ak}, and L_{Bk}). Hence, a fortiori, Lemmas 11, 12, and 14 continue to hold.

Two lemmas remain to be checked: Lemma 6 and Lemma 13. Lemma 6 as well as Equation (17) now hold for some (b + K) $\times$ a matrix Λ. Also, the matrices Z_B and M_B are now defined as follows:

$$Z_B = \begin{bmatrix} Z_{B1} \\ \cdot \\ \cdot \\ Z_{BK} \end{bmatrix} \quad ; \quad M_B = \left[N_C \ ;\ Z_B\right] = \left[e^1_K \otimes U_J, \ldots, e^K_K \otimes U_J \ ;\ Z_B\right] \tag{A.93}$$

The proof of this modified Lemma 6' is similar to the proof of Lemma 6, and relies on the equation

$$E = -\frac{\partial \beta' Z_B'}{\partial \alpha}\left[V_{B|C} - V_{B|C}Z_B\left[Z_B'V_{B|C}Z_B\right]^{-1}Z_B'V_{B|C}\right]\frac{\partial Z_B \beta}{\partial \alpha}$$

(A.94)

where $V_{B|C}$ is the block diagonal matrix of which the k-th block is $p_k \Omega_{B|k}$. Equation (A.94) straightforwardly follows from Equations (17), (36), and Lemma 3 which now holds for L_{Bk}.

It then follows that Corollary 1 is modified accordingly. Furthermore, the proof of Corollary 2 shows that this corollary still holds provided the model space $M_B(\alpha)$ be now defined as the linear vector space spanned by the column vectors of the matrix M_B defined in (A.93).

Finally, in order to extend Theorem 3 to the case of qualitative individual characteristics, we need to show that Lemma 13 still holds where u and w are defined as follows:

$$u = \sum_k p_k u_k \quad ; \quad w = \sum_k p_k w_k \tag{A.95}$$

where

$$u_k = \begin{bmatrix} \dfrac{\partial^2 L_{Ak}}{\partial\alpha\partial f'_{AB|k}} + \dfrac{\partial^2 L_{Bk}}{\partial\alpha\partial f'_{AB|k}} \\ \\ \dfrac{\partial^2 L_{Bk}}{\partial\beta\partial f'_{AB|k}} \end{bmatrix} (f_{AB|k} - Pr^o(A,B|k)) \tag{A.96}$$

and

$$w_k = \left[-G \frac{\partial^2 L_{Ak}}{\partial\alpha\partial f'_{AB|k}} + \frac{\partial^2 L_{Bk}}{\partial\alpha\partial f'_{AB|k}} - \frac{\partial^2 L_B}{\partial\alpha\partial\beta'} \left[\frac{\partial^2 L_B}{\partial\beta\partial\beta'}\right]^{-1} \frac{\partial^2 L_{Bk}}{\partial\beta\partial f'_{AB|k}} \right] (f_{AB|k} - Pr^o(A, B|k)) \tag{A.97}$$

Compare (A.94) and (A.97) to (A.60) and (A.65). Note that (A.97) also involves G and L_B. Then it can readily be shown using Lemmas 1' and 5 that Lemma 13 holds with the above definitions. It then follows that the proof of Theorem 3 goes through provided one define v as:

$$v = \sum_k p_k v_k$$

where

$$v_k = \begin{bmatrix} \dfrac{\partial^2 L_{Ak}}{\partial\alpha\partial f'_{AB|k}} \\ \dfrac{\partial^2 L_{Bk}}{\partial\beta\partial f'_{AB|k}} \end{bmatrix} (f_{AB|k} - Pr^o(A,B|k))$$

(See (A.60)).

Footnotes

1. I am much grateful to David Grether for his comments and encouragement. I have also benefited from stimulating discussions with John Link and Marc Nerlove.

2. Thus I and J are constant. Since the model (1) - (5) requires all the probabilities to be strictly positive, this is a simplification. The results of this paper can, however, be straightforwardly extended to cases in which I depends on t and on the value of B, and J depends on t. Also, to simplify the notations, the subscript "t" has been suppressed from the random variables A_t and B_t.

3. In Q. H. Vuong (1982b), examples of such functions $z_{jt}(.)$'s are given. One may want to consider models in which the functions $z_{jt}(.)$'s are known only up to some unknown parameters θ. However, one would then lose the convenient feature that the probability model for B is log-linear or logit.

4. In particular, by making such an assumption, we avoid the problems associated with cases in which the set of possible values of the explanatory variables in (1) - (2) is unbounded due, for instance, to undesigned experiments. We shall return to this point in Section 5.

5. It is more convenient to order the components of log Pr(A|B) and the rows of Z_A according to the inverse rather than to the usual

lexicographical ordering of the pairs (i,j).

6. As a matter of fact, the recursive pair of probability models (3.1) - (3.2) is defined by

   (i) a conditional log-linear probability (CLLP) model for A given B with model space generated by the J vectors ν^j and the a column-vectors of Z_A,

   (ii) a log-linear probability model for B with a model space generated by U_J and the b column-vectors of Z_B, and hence depending on a.

   (See Q. H. Vuong (1982a)).

7. It is well known that the non-existence of M. L. estimate in log-linear probability models or logit models is due to the presence of observed empty cells. For a characterization of the cases in which the M.L. estimate exists, see, e.g., S. J. Haberman (1974), D. McFadden (1974), and J. P. Link (1983).

8. The covariance matrix $\sum^1$ can clearly be consistently estimated by the right-hand side of (14) where all the second partial derivatives of L_A and L_B are evaluated at $(\hat{\alpha}^1, \hat{\beta}^1, f_{AB})$. Let us note that other consistent estimators of $\sum^1$ exists. For instance, one may use the right-hand side of (14) where all the partial derivatives are evaluated at $(\hat{\alpha}^1, \hat{\beta}^1, \hat{Pr}^1(A,B))$.

9. Throughout this paper, the asymptotic covariance matrix of a consistent estimator $\hat{\theta}$ is the covariance matrix of the asymptotic distribution of $\sqrt{T}(\hat{\theta} - \theta^o)$ as T goes to infinity.

10. The covariance matrix $\sum^M$ is nonsingular as shown in the Appendix (see the discussion following (A.60)).

11. A consistent estimator $\hat{\theta}^2$ with asymptotic covariance matrix $\sum^2$ is said to be less efficient than another consistent estimator $\hat{\theta}^1$ with asymptotic covariance matrix $\sum^1$ if $\sum^2 - \sum^1$ is a positive semi-definite matrix.

12. The conditional frequencies $f_{i|j}$ are all well-defined since we have assumed that there are no empty cells. We shall return to this assumption in Section 5.

13. Note that in (ii) we use f_B instead of $\hat{f}_B^r = \hat{Pr}^r(B)$. If, however, $\hat{f}_B^r$ is used, then it can be proven that Theorem 2 continues to hold.

14. I am much indebted to J. P. Link for illuminating discussions on this point.

15. Moreover, the negative of the inverse of the second partial derivatives of L_A evaluated at $(\hat{\alpha}^1, f_{AB})$ is a consistent estimator of the asymptotic covariance matrix of the two-step conditional estimator of α. Hence this matrix is usually given by standard computer packages for logit model estimation.

16. The effect of using $\hat{f}_B^r = \hat{Pr}(B)$ instead of f_B in (ii) will be studied in future work.

17. As in footnote 8, the asymptotic covariance matrix $\sum^r$ can be consistently estimated by the right-hand side of (28) where all the partial derivatives are evaluated at $(\hat{\alpha}^r, \hat{\beta}^r, f_{AB})$ or $(\hat{\alpha}^r, \hat{\beta}^r, \hat{Pr}^r(A, B))$.

18. It is easy to see that the asymptotic covariance matrices of the $\hat{\delta}^r$'s are all equal if and only if $G^2 = I$. From Lemma 14 of the Appendix, it follows that this holds if and only if all the roots of G are equal to one, i.e., G is similar to the identity matrix.

19. As a matter of fact, the adjustment (24) defining the efficient iterative sequential procedure was discovered by noting the property stated in Corollary 6.

20. See Q. H. Vuong (1928a, 1982b). This is the classical identification problem as defined by e.g., T. J. Rothenberg (1971).

21. For general theorems on the existence of M.L. estimates in log-linear probability models, see S. J. Haberman (1974). For straightforward extensions of these results to CLLP models, see Q. H. Vuong (1982a). See also J. P. Link (1983).

22. If C is non-trivially polytomous, then K is simply the product of the numbers of categories of all the qualitative individual characteristics.

23. It can easily be shown that all the parameters in α are identified if and only if the matrix $Z_A^* = [Z_{Ak}^{*\prime}$; k observed$]'$ is full column-rank where Z_{Ak}^* is the matrix obtained from Z_{Ak} by deleting the rows associated with unobserved combinations (j, k). Similarly, all the parameters in β are identified if and only if the matrix $Z_B^* = [Z_{Bk}'$; k observed$]'$ is full column-rank. (See Q. H. Vuong (1982a)).

24. The adjustment (20) which is associated with the natural iterative sequential procedure is generalized in the same way.

25. We are considering exogenous sampling only. If $Pr^o(C)$ is equal to the true marginal distribution of the characteristic C in the population, then one has a random exogenous sampling; otherwise, one has a stratified exogenous sampling (see C. Manski and D. McFadden (1981) for this terminology and other sampling designs).

26. Specifically, the adjustments (23) and (24) associated with the natural and efficient iterative sequential procedures now apply to the conditional frequency $\hat{f}^r_{AB|k}$ (as in Equation (39)). Then, the statements of Theorem 1 - 3, and Corollaries 1 - 6 stay as they are (the matrices E, F, G, are defined as in Sections 2 - 4, and L_A and L_B (not L_{Ak} and L_{Bk}) are used in the formulas), the only exception being Corollaries 1 and 2, which are modified as discussed in Section 5 of the Appendix.

27. More precisely, we now use Z_{Ak} and $V_{A|Bk}$ instead of Z_A and $V_{A|B}$ in (A.11) where $V_{A|Bk}$ is defined as in (A.8) - (A.10) with Pr(A|B, k) instead of Pr(A|B). Also, a consistent estimate of the asymptotic covariance matrix $\sum^r$ can be obtained by evaluating all the partial derivatives in the right-hand side of (28) at $(\hat{\alpha}^r, \hat{\beta}^r, \hat{Pr}^t(A,B|C), f_C)$ or $(\hat{\alpha}^r, \hat{\beta}^r, f_{AB|C}, f_C)$.

28. These technicalities essentially arise from the fact that we do not have many observations per cell, i.e., repeated observations.

29. Its asymptotic covariance matrix is given by (14). Actually, the asymptotic covariance matrix is the probability limit of the matrix in the right hand side of (14) where all the second partial derivatives are evaluated at (α, β, y_{ABX}).

30. When there are no repeated observations the method that consists in substituting $f_{AB|t}$ and $f_{B|t}$ for Pr(A|B) and Pr(B) in Equation (A.8) - (A.9) leads to an estimated matrix $\hat{V}_{A|B}$ that is identically null.

31. See footnote 29.

REFERENCES

Amemiya, T.: "On a Two-Step Estimation of a Multivariate Logit Model," Journal of Econometrics, 8 (1978), 13-22.

Andersen, E. B.: "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," Journal of the Royal Statistical Society, Series B, 32 (1970), 283-301.

Berndt, E. R., Hall, B. H., Hall, R. E., and Hausman, J. A.: "Estimation and Inference in Non-Linear Structural Models," Annals of Economic and Social Measurement, 3 (1974), 653-665.

Domencich, T. A., and McFadden, D.: Urban Travel Demand. Amsterdam: North Holland, 1975.

Gantmacher, F. R.: The Theory of Matrices, Volume One. New York, N. Y.: Chelsea Publishing Company, 1960.

Haberman, S. J.: The Analysis of Frequency Data. Chicago: University of Chicago Press, 1974.

Link, J. P.: "Estimativity in Log-Linear Probability Models," forthcoming PhD Thesis, Department of Economics, Northwestern University, 1983.

Mann, H. B., and Wald, A.: "On Stochastic Limit and Order Relationships," Annals of Mathematical Statistics, 14

(1943), 217-226.

Manski, C., and McFadden, D.: "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in Structural Analysis of Discrete Data with Econometric Applications. Ed. by C. Manski and D. McFadden. Cambridge, Mass.: MIT Press, 1981.

McFadden, D.: "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers of Econometrics. Ed. by P. Zarembka. New York: Academic Press 1974.

______: "Econometric Models of Probabilistic Choice," in Structural Analysis of Discrete Data with Econometric Applications. Ed. by C. Manski and D. McFadden. Cambridge, Mass: MIT Press, 1981.

Nerlove, M., and Press, S. J.: Univariate and Multivariate Log-Linear and Logistic Models, Santa Monica: Rand Corporation, 1973.

______: "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data," Discussion Paper Number 1, Center for Statistics and Probability, Northwestern University, 1976.

Rao, C. R.: Linear Statistical Inference and its Applications. New York, N. Y.: Wiley, 1965.

Rothenberg, T. J.: "Identification in Parametric Models," *Econometrica*, 39 (1971), 577-592.

Rothenberg, T. J., and Leenders, C. T.: "Efficient Estimation of Simultaneous Equation Systems," *Econometrica*, 32 (1964), 57-76.

Vuong, Q. H.: "Conditional Log-Linear Probability Models: A Theoretical Development with an Empirical Application," unpublished Ph.D. Thesis, Department of Economics, Northwestern University, 1982.

_______________: "Probability Feedback in a Recursive System of Probability models," unpublished manuscript, California Institute of Technology, 1982.